

# 통계학 교육용 한글 소프트웨어 개발 연구<sup>†</sup>

이정진\* · 강근석\* · 이윤오\*

## 요 약

통계학을 배우는 초보자들을 위하여, 퍼스날 컴퓨터를 이용한 통계교육용 소프트웨어 (Computer Aided Teaching for Statistics : CATS)를 한글을 사용하여 개발하였다. 이 소프트웨어는 다량의 자료를 처리하기 보다는, 소량의 통계자료를 가지고 기초 통계학의 여러 기법을 실습하거나, 초보자가 책만 가지고 이해하기 힘든 이론을 컴퓨터를 이용하여 교육시키는데 주목적이 있다.

## 1. 서 론

통계학은 사회과학, 자연과학, 공학등 여러 학문분야에서 그 중요성이 인식되어, 통계학을 배우려고 하는 사람의 수가 점차로 늘어나고 있다. 하지만 통계학을 배우는 많은 초보자(특히 인문, 사회적 배경을 가진 사람)들이 통계학을 배우는 것이 힘들다고 불평한다. 여러 가지 이유가 있을 수 있겠으나, 그 첫번째 이유는 통계학이란 학문의 성격상 수학을 많이 이용하기 때문에 배우기가 어렵다는 것이다. 두번째 이유는 초보자에게 어렵게 쓰여진 통계학책에 문제가 있을 수 있다고 생각된다. 이밖에도 필자들의 경험으로는 강의나 책으로 설명하기가 힘든 이론이 통계학에는 너무 많다. 따라서 초보자들에게 “어떻게 쉽고 올바르게 통계학 교육을 효율적으로 시킬 수 있는가?”라는 문제는 모든 통계학자들의 공통 관심사중의 하나이다.

20세기 문명의 총아인 컴퓨터는 위 문제의 해결에도 도움을 줄 수 있다고 생각된다. 현재 SAS, SPSS, BMDP등 좋은 통계 팩키취가 개발되어 있다. 하지만 기존의 컴퓨터 통계 팩키취는 초보자 교육용이라기보다는, 이미 통계학을 알고 있는 사람들이 다량의 자료를 전산처리하는데 더 적합하다. 또한 이들 통계 팩키취들은 영어로 쓰여져 있어서, 대부분이 한글로 쓰여진 통계학 책으로 교육받는 초보자들에게는 그 용어들이 여간 생소하지가 않다.

---

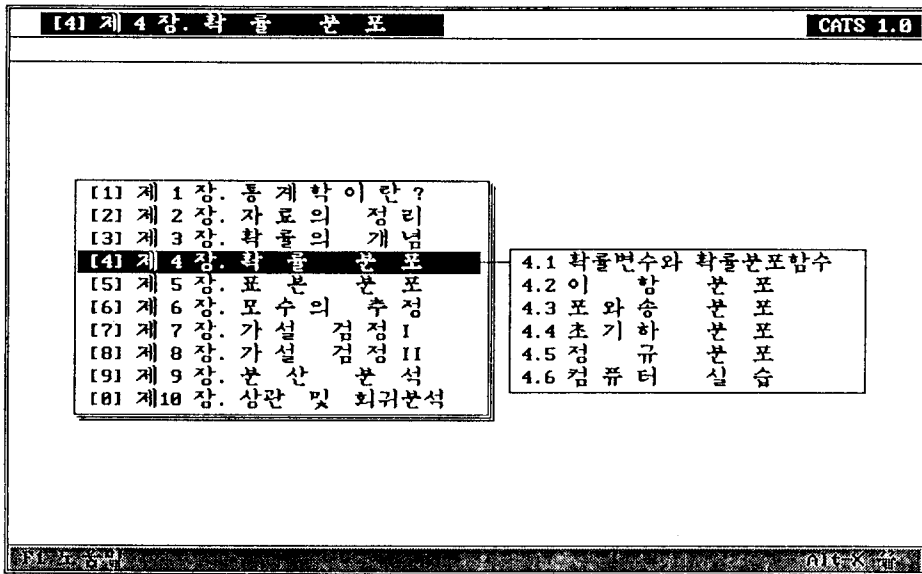
† 이 논문은 1990년도 문교부 지원 한국학술진흥재단의 자유공모과제 학술연구조성비에 의해 연구되었음.

\* (156-743) 서울특별시 동작구 상도동 1-1 송실대학교 통계학과

본 연구에서는 통계학을 배우는 초보자들을 위하여, 퍼스날 컴퓨터를 이용한 통계교육용 소프트웨어 (Computer Aided Teaching for Statistics : CATS)를 한글을 사용하여 개발하였다. 이 소프트웨어는 다량의 자료를 처리하기 보다는, 소량의 통계자료를 가지고 통계학의 여러 기법을 실습하거나, 초보자가 책만 가지고 이해하기 힘든 이론을 컴퓨터를 이용하여 교육시키는데 주목적이 있다.

## 2. 소프트웨어 개요

본 연구에서 개발한 통계교육용 한글 소프트웨어는, <그림 1>과 같이 대부분의 기초통계학 책에서 다루고 있는 아래의 통계 기법을 퍼스날 컴퓨터를 이용하여 교육하고, 또 소량의 자료를 가지고 실습할 수 있도록 고안되었다.



<그림 1> 풀다운메뉴 형식의 시스템

각 장은 모두 두 부분으로 되어있다. 첫 부분은 기존의 통계학 책의 내용을 컴퓨터 스크린에 간추려 놓은 것이고, 두번째 부분은 첫 부분에서 배운 내용을 컴퓨터를 이용하여 실습하는 것이다. 본 논문은 이 두번째 부분의 중요한 내용을 일부 발췌한 것이다. 전체 시스템은 풀다운 메뉴 형식으로 고안되었고, 입력자료를 받는 프로그램은 로터스 스프레드시트(LOTUS spreadsheets) 시스템과 비슷하게 설계되어 사용자가 편리하게 설계되었다.

이 소프트웨어는 Borland사의 Turbo C (Ver 2.0) 컴퓨터언어를 사용하여 개발 되었는데 대략 20,000 줄의 C 언어 문장으로 이루어져 있다. IBM/PC의 호환기종에 허큘리스 그래픽스 카드(Hercules graphics card)가 장착되어 있는 모노크롬 모니터가 있으면 설치가 가능하다. 두 개의 플러피

시스템에서도 가능하지만 하드 디스크가 있어야 매우 편리하다. 필요한 디스크 저장용량은 약 3메가바이트이고, 주기억장치는 640K 바이트 이상이 필요하다. 이 소프트웨어는 한글을 사용하지만 자체내의 한글사용루틴을 쓰기 때문에 한글보드가 불필요하여 기존의 시장에 나와 있는 어떠한 국산 퍼스날 컴퓨터에서도 사용이 가능하다.

### 3. 소프트웨어 주요내용

본 논문에서는 지면관계상 개발된 소프트웨어의 내용중 기존의 통계학책에서 설명하기 힘들었던 이론이나, 통계팩키지에서 다루기는 했어도 설명이 불충분하여 이해하기 어려웠던 부분만을 발췌하였다. 직접 예제자료를 입력하여 컴퓨터실행을 하여 소개하고자 한다.

#### 3.1 자료의 입력

〈그림 2〉는 개발된 소프트웨어에서 사용되는 자료입력 부분이다. 스프레드시트(spreadsheet) 형식으로 디자인 하여 커서를 사용하면서 필요한 숫자를 입력하면 된다. 여기서 행은 관찰대상 번호, 열은 변수를 나타내는데 편의상 각 변수의 관찰대상의 수가 같다고 가정하였다. 현재는 기존화일 불러들이기, 숫자 입력, 화일저장 기능만 가지고 있는데 차후 자료의 변환도 가능하게 일반화 시키려고 한다.

자료를 각 셀에 입력 하십시오.    행: 관찰자번호, 열: 변수번호  
 변수 1    변수 2    변수 3    변수 4    변수 5    변수 6    변수 7

1	1	1				
2	2	2				
3	3	3				
4	4	1				
5	5	2				
6	6	3				
7	7	1				
8	8	2				
9	9	3				
10	10	2				
11						
12						
13						
14						
15						
16						
17						
18						
19						

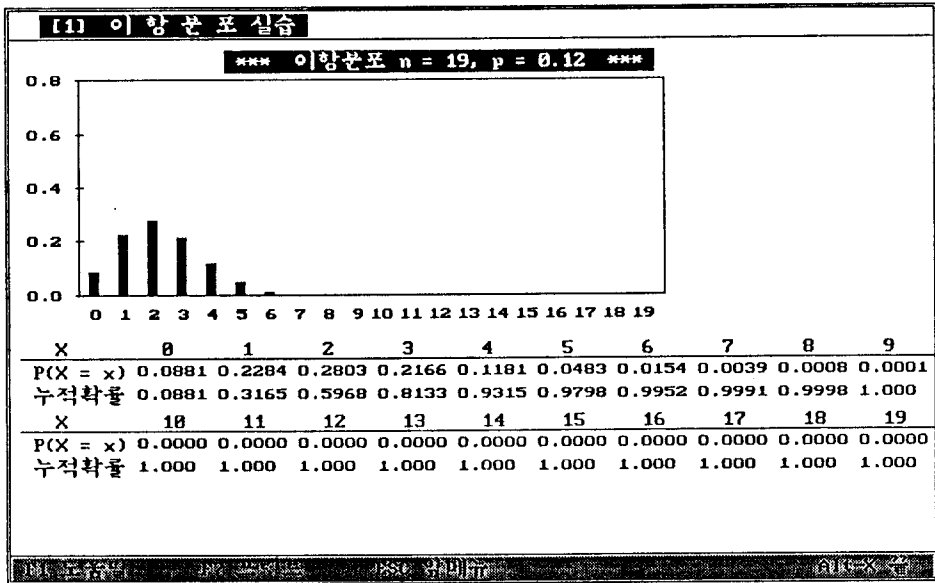
〈그림 2〉 자료입력 부분

### 3.2 기술통계

입력한 자료를 이용하여 도수분포표, 히스토그램, 측도의 계산, 줄기-잎 그림(stem-leaf plot) 원형차트(pie chart)등의 간단한 자료처리를 할 수 있다.

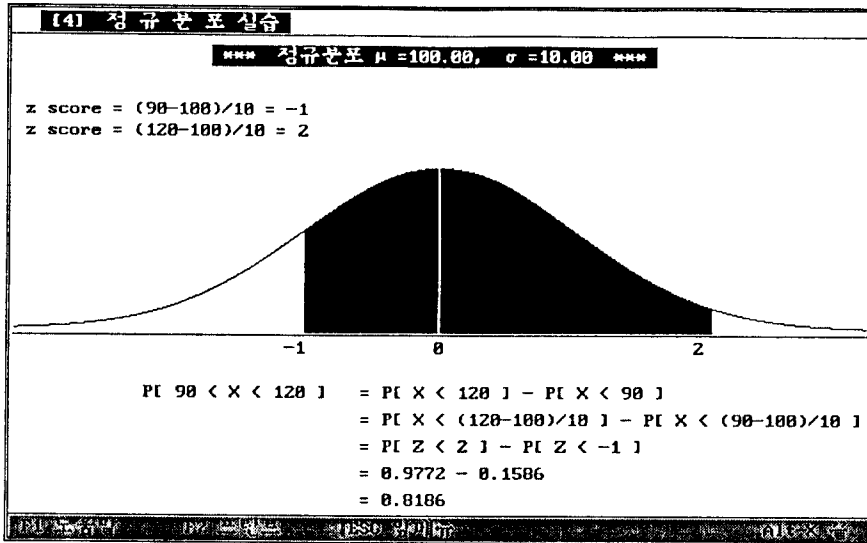
### 3.3 확률분포

이항분포, 포와송분포는 대개의 기초통계책에서 소개되는 확률분포이고, 여러 경우의 파라메타에 대한 확률값을 부록에서 소개하는데 지면관계상 많은 파라메타값을 취급하지는 못한다. 예를 들면 ([2]) 이항분포의 경우  $n$ 의 값은 2에서 25까지 다루고,  $p$ 의 값은 0.05 0.10 0.20 ... 0.90 0.95 까지 간격 0.05나 0.10에 대하여 이항확률값을 표로서 보여준다. 개발된 소프트웨어에서는 이항분포의 경우 사용자가  $n$ (스크린 크기의 제한으로 29까지만 허용)과  $p$ (0과 1 사이의 임의의 실수)를 입력하면 분포표의 그림과 그 확률을 스크린에 보여준다. 따라서 이 소프트웨어를 이용하면 <그림 3>과 같이 여러 가지 다양한 모양의 이항분포를 연구할 수 있다. 이외에 포와송분포, 초기하분포도 실습이 가능하다.



<그림 3> 이항분포 실험

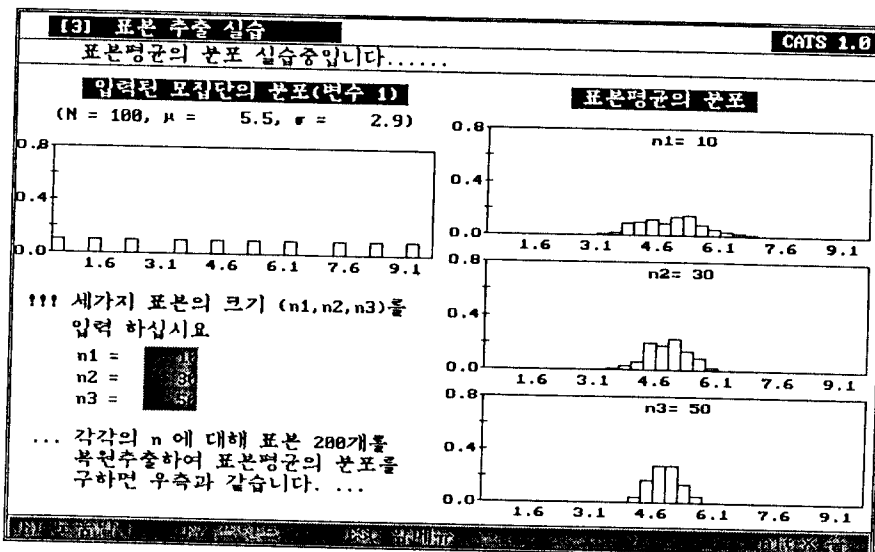
정규분포의 경우에는 대개의 책에서 표준정규변수  $Z$ 의 누적확률  $P(Z < z)$ 을  $z$ 가 -3.00 -2.99 ... 3.00에 대하여 표로서 보여주고 있다. 본 소프트웨어에서는 표준정규분포는 물론이고 평균이  $\mu$ , 분산이  $\sigma^2$ 인 임의의 정규분포에 대해서도 <그림 4>와 같이 세 가지 확률형태 ( $P(X < x)$ ,  $P(a < X < b)$ ,  $P(X > x)$ )에 대해 정규화 과정을 보여주며 그 확률을 나타내는 그림과 계산된 확률을 보여준다.



<그림 4> 정규분포 실습

### 3.4 표본분포

통계학의 이론중 가장 초보자들이 이해하기 힘든 것 중의 하나가 필자들의 교육경험에 의하면 표본분포라고 생각된다. 이 표본분포는 기존의 어느 통계팩키지를 이용하더라도 실습하기가 용이하지 않다. 본 소프트웨어에서는 모집단 자료를 입력한 후 표본의 크기 세 가지에 대해 200개의 서로 다른

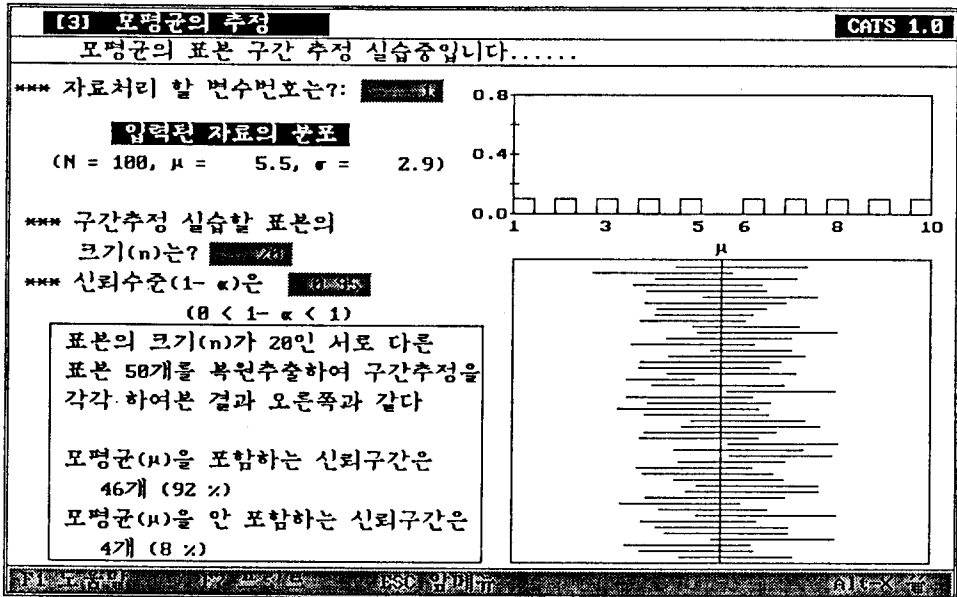


<그림 5> 표본평균의 분포

표본을 복원추출하였을 때 어떻게 표본분포가 달라지는 가를 실습할 수 있다. <그림 5>는 표본평균의 분포로서 모든 가능한 표본은 아니더라도 중심극한 정리가 무엇인지 잘 보여주고 있다. 마찬가지로 표본분산의 분포와 표본비율의 분포도 실습할 수 있다.

### 3.5 모수의 추정

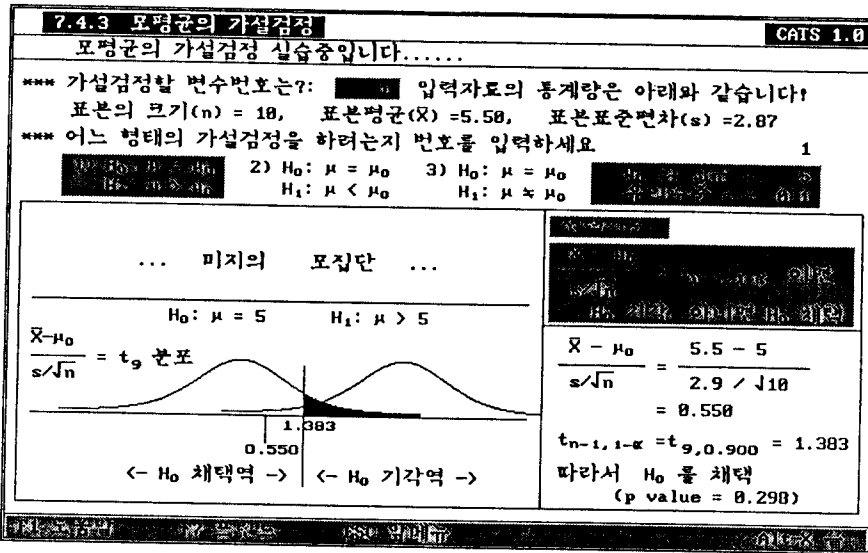
모평균에 대한 95% 신뢰도 구간추정의 의미를 설명하는 것도 쉬운일이 아니다. 본 소프트웨어에서는 모집단 자료를 입력한 후 정해진 표본의 크기와 신뢰계수에 대해 50개의 구간추정을 스크린에 보여 주면서, 이 중 몇%가 모수를 포함하는지 알려준다. <그림 6>은 모평균에 대한 구간추정실습이다. 이 외에 모분산과 모비율의 구간추정도 실습할 수 있다.



<그림 6> 모평균의 구간추정실습

### 3.6 가설검정

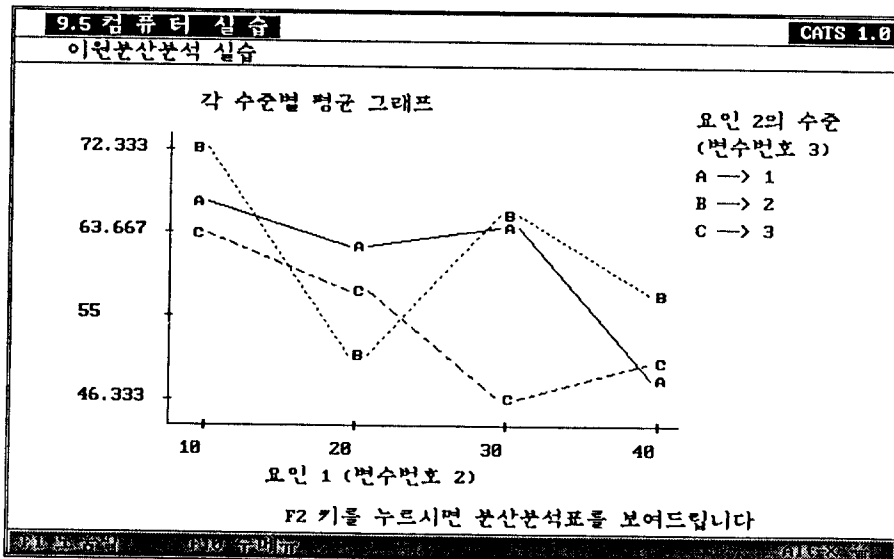
가설검정의 이해를 돕기 위해서 두 미지의 가설에 대한 표본분포를 스크린에 그린 다음 유의수준에 따른 기각역을 나타내어 준다. 그리고 입력된 표본자료에서 산출된 표본통계량이 이 표본분포의 어느 부분에 위치하여 있는 지 보여 주고, 선택기준에 따라 귀무가설의 기각여부를 알려준다. <그림 7>은 모평균에 대한 가설검정이다. 컴퓨터를 이용하면 모분산을 모르는 경우 표본의 크기가 크더라도 정규분포 근사를 쓰지않고 t 분포를 사용할 수 있다. 마찬가지로 모분산과 모비율에 대한 가설검정을 실습할 수 있다.



〈그림 7〉 모평균의 가설검정실습

### 3.7 분산분석

이 부분에서는 일원분산분석, 이원분산분석과 난괴법(randomized block design)을 설명하고 있다. 일원분산분석에서는 분산분석표, 각 수준별 평균, 표준편차, 그리고 분산의 합동추정치(pooled va-



〈그림 8〉 이원 분산분석에서의 평균그래프

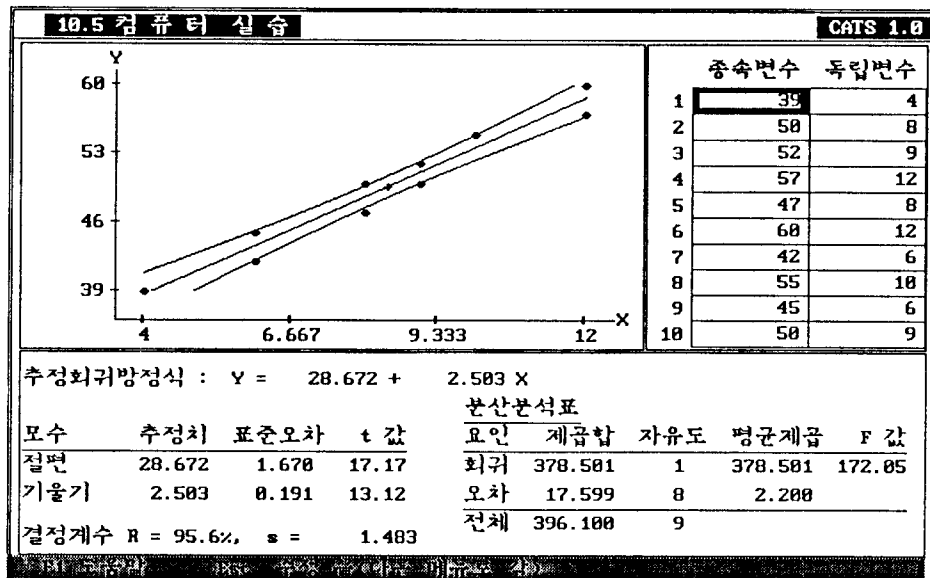
riance)를 사용한 95% 신뢰구간을 보여준다. 이 외에 관측자료의 산포도를 볼 수 있다. 난괴법은 일원분산분석의 처리결과와 비슷하다. 이원분산분석에서는 분산분석표와 각 수준조합별 평균, 표준편차표 외에 이원분산분석을 설명하는데 중요하면서도 기존의 통계패키지들이 보여주지 못하는 각 수준별 평균그래프를 보여준다(그림 8). 교호작용의 의미등을 이 평균그래프를 사용하여 쉽게 설명할 수 있을 것이다.

### 3.8 회귀분석

〈그림 9〉는 회귀분석의 기본출력양상을 보여준다. 처음에는 산포도와 관측자료, 분산분석표 등의 결과를 보여주며, 이후에 적합된 회귀직선, 신뢰대를 선택적으로 보이게 할 수 있다. 그리고, 이 소프트웨어의 장점이라 할 수 있는 자료수정과 재실행이 가능하다. 즉, 관측자료화면에서 원하는 자료를 수정한 다음 재실행 키를 누르면 수정된 자료에 대한 회귀분석결과를 보여준다. 이 기능을 이용하면 회귀분석에 대한 각 자료의 영향(influence) 정도를 시각적으로 보여줄 수 있을 것이다.

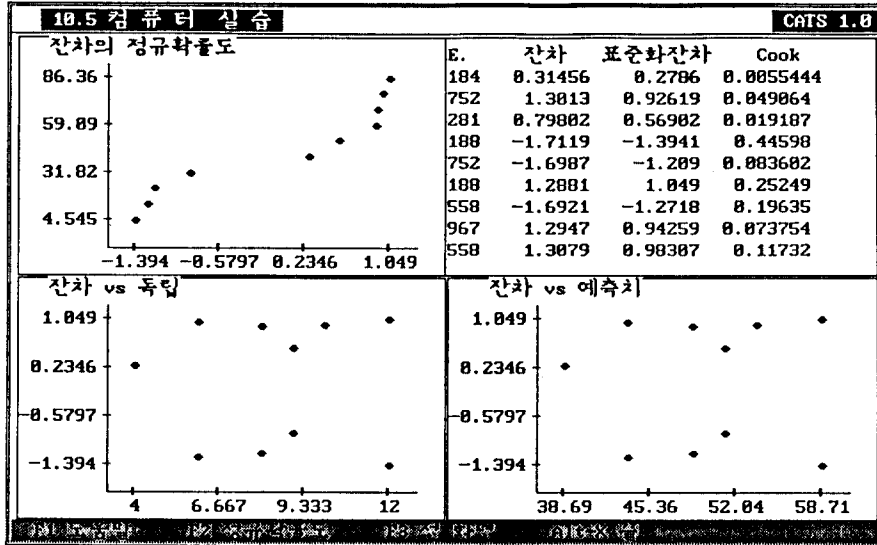
〈그림 10〉은 잔차분석을 위한 화면으로 관측자료, 잔차, 표준화잔차, Cook의 영향측도등과 함께 잔차의 독립변수와 예측치에 대한 산포도를 보여주고, 잔차의 정규확률도를 선택적으로 보여준다. 〈그림 9〉와 〈그림 10〉은 언제든지 번갈아 가면서 볼 수 있게끔 작성되어 있다.

본 소프트웨어의 회귀분석을 위한 프로그램은 독립변수의 수에 관계없이 이용될 수 있으나, 현재로는 단순선형회귀만 가능하고 중선형회귀는 출력양식에 대한 의견을 모아 곧 첨가할 예정이다.



〈그림 9〉 단순 선형회귀분석





〈그림 10〉 잔차분석

#### 4. 향후과제 및 결론

본 논문에서 소개한 통계학 교육용 한글 소프트웨어는 현재 거의 마무리 되어 '92년 1학기부터 숭실대 통계학과 2학년 기초통계 교육에 시험적으로 사용되고 있다. 현재는 Epson 도트프린터만 지원하는데 앞으로 여러 종류의 프린터를 지원할 수 있고, 또 여러 종류의 그래픽스 보드(graphics board)와 모니터를 지원하도록 하는 보완작업을 거쳐 일반 사용자들에게 배포될 예정이다. 그 때가 되면 이 소프트웨어의 유용성에 관해서 많은 통계학자들이나 사용하는 학생으로부터 좋은 충고와 지적을 받으리라 생각하고 있다. 그리고 미래의 과제로서 남은 기초통계분야를 보강하고, 또 고급 통계분야도 각 전공분야의 학자들과 함께 개발하였으면 한다.

이 소프트웨어 개발의 가장 큰 의의는 컴퓨터를 이용한 새로운, 또 다양한 통계교육의 가능성을 보여 주었다는데 있다고 생각한다. 이 소프트웨어를 이용하여 통계학을 처음으로 배우는 사람들이 책만 가지고는 이해하기 힘든 이론을 쉽게 이해할 수 있게 되고, 컴퓨터를 잘 몰라도 기초통계 수준의 자료처리를 할 수 있게 되기를 기대한다.

이 소프트웨어 개발의 다른 의의는 한글 통계팩키지를 개발할 수 있는 여러 가지 기술축적에 있다. 필자가 알기로 우리나라 기관이나 기업에서 일년에 SAS나 SPSS 입대에 들어가는 비용이 몇 백만불 소요된다고 한다. 실제 이들이 사용하는 통계기법의 대부분은 별로 복잡한 것이 아니라고 한다. 현재 한국 통계학회 회원은 대략 500명이 넘는다고 하는데, 여러 가지 이유가 있을 수 있지만 이렇게 수

많은 통계인이 있으면서 우리의 한글 통계팩키지가 하나도 없다는 것은 어떻게 보면 수치스럽게 느껴지기도 한다. 이 소프트웨어를 개발하면서 SAS나 SPSS에 미치지 못하는 우리의 한글 통계팩키지를 개발할 수 있는 상당한 기술 축적을 하였다고 볼 수 있다. 조만간 중지를 모아 한글 통계팩키지 개발이 실현되기를 고대한다.

## 참 고 문 헌

- [1] 한국통계학회(1987), 통계용어사전, 자유아카데미.
- [2] 김우철외 7인(1990), 현대통계학, 제3판, 영지문화사.
- [3] Daniel, W.W. and Terrell, J.C.(1989), *Business Statistics*, 4th Edition, Houghton Mifflin.
- [4] Chou, Ya-lun (1980), *Statistical Analysis for Business and Economics*, Elsevier.
- [5] McClave, J.T. and Benson, P.G. (1988), *Statistics*, 4th Edition, Dellen/Macmillan.
- [6] Schildt H. (1990), *Turbo C The Complete Reference*, Boland/McGraw-Hill.

## A Study of Computer Aided Teaching for Statistics<sup>†</sup>

Jung Jin Lee<sup>\*</sup>, Gunseog Kang<sup>\*</sup> and Yoon Oh Lee<sup>\*</sup>

### ABSTRACT

A software, the CATS (Computer Aided Teaching for Statistics), which is specially designed for teaching elementary statistics using the personal computer is developed. This software helps the novice to understand the complicated statistical theories, for example the central limit theorem, which are not easy to teach in classroom. The software is also designed for practicing the elementary statistical techniques using the small amount of data as the other statistical packages.

---

<sup>†</sup> This paper was supported in part by the NON DIRECTED RESEARCH FUND, Korea Research Foundation, 1990.

<sup>\*</sup> Department of Statistics, Soongsil University, Sangdo-Dong 1-1, Dongjak-Ku, Seoul, Korea (156-743)