

New Paradigm in Statistical Education

Jae-Woo LEE¹⁾, Jea-Young LEE²⁾

Abstract

Korea has the most level of Internet Infrastructure in the world. But, in the educational aspect, it does not have an enough foundation about Statistical Education. In this paper we consider the methods of activation about statistics. Also, we present what is the Enterprise Guide and what does it have characteristics as statistical analysis tool from educational point of view. And we suggest a new paradigm in statistical education.

Keywords : 인터넷, Enterprise Guide, Thin-Client

1. 서론

정보통신부가 2002년 한국 인터넷 통계집에 발표한 자료에 따르면 2002년 6월 현재 인터넷의 국내 인터넷 이용자 수가 2,565만 명(이용률 58%)을 보였고, 직업별 인터넷 이용률은 학생이 무려 94.5%의 이용률을 보이고, 대학생의 인터넷 이용률은 97.7%에 이른다. 이는 한국의 정보통신 기술의 발달에 따른 당연한 결과라고 판단되며 이러한 추세로 볼 때 전 국민의 인터넷 이용도 머지않았다고 할 수 있다. 이러한 가운데 인터넷을 통한 정보 접근과 활용은 우리 실생활에서 필수 불가결한 요소로 자리잡았으며, 이는 앞으로의 교육체제도 변해나가야 함을 뜻한다. 이미 많은 분야에서는 인터넷을 통한 온라인 교육이 시행되고 있으며 매우 활성화되어 있다. 하지만 통계분야에서는 일반 온라인교육과는 달리 실제 데이터의 분석이라는 작업이 수반되어야 하므로 그리 용이하지 않았다. 본 연구에서는 통계 교육의 현재를 2장에서 살펴보고, 3장에서 온라인을 이용한 thin-client 통계분석 소프트웨어를 소개하며 4장에서 앞으로의 통계 교육의 나아갈 길을 제시하고자 한다.

2. 통계교육의 현재

예전의 통계교육에서는 교재에 실려있는 데이터, 즉 소규모의 자료를 직접 입력해서 데이터를 만들고 그것을 이용하여 분석하는 형태였다. 그러다가 엑셀 형태의 스프레드시트나 데이터베이스 형태의 MS Access와 같은 데이터를 읽어들이 분석하는 GUI 환경의 소프트웨어들이 제공되었다. 하지만 현재 분석되는 데이터들은 시스템의 발달과 분석용구의 변화(예: 데이터마이닝)에 따른 대용량의 데이터이며, 이는 현재 통계 교육에서도 무시될 수 없는 조건이 되었다. 또한 GUI환경의 사용하기 쉬운 소프트웨

1) 서울시 영등포구 여의도동 알리안츠타워 SAS KOREA IA 팀장 E-mail : jae-woo.lee@sas.com

2). 경북 경산시 대동 214-1 영남대학교 이과대학 수확통계학부 부교수

어릴지라도 분석작업에 대한 수행과정을 공유하고자 하면 반드시 코드를 이용하여야만 하였다. 이럴 경우, 해당 소프트웨어에 익숙하지 않은 사용자는 그 코드를 아무런 해석없이 그대로 따라서 실행해야 하거나 그와 똑 같은 작업을 GUI환경에서 하기 위해서는 결국 해당 소프트웨어의 코드를 전부 익혀야만 하였다. 이는 학생들로 하여금 통계소프트웨어와의 거리를 생기게 만들며 결국에는 통계 자체에 대한 흥미를 줄어 들게 만드는 계기가 된다. 물론 이론적인 연구가 소홀해 저서는 안되지만 실제 사회에서 통용될 수 있는 분석능력을 갖게 하는 것도 학원이 해야 할 의무중에 하나이기에 학생들에게 좀 더 친숙하고 쉽게 다룰 수 있는 통계소프트웨어를 활용하게 하는 것은 피할 수 없는 선택이다. 이러한 점들을 고려해 볼 때 통계분석에 필요한 소프트웨어는 학생들에게 쉽게 다가설 수 있는 인터페이스를 가졌으며 또한 인프라가 충분한 곳에서는 막강한 분석능력을 그대로 이용할 수 있어야 한다.

2002년 여름에 발표된 자료에 따르면[표 1 참조] 우리나라는 현재 인터넷에 있어서 세계 제일의 위치에 있다고 할 수 있다. 또한 2002년 한국 인터넷 통계집에 따르면 대한민국의 대학생 거의가 인터넷을 활용하고 있음을 알 수 있으며 이는 온라인 통계 교육을 위한 인프라가 충분히 갖춰져 있다고 볼 수 있다.[그림1 참조]

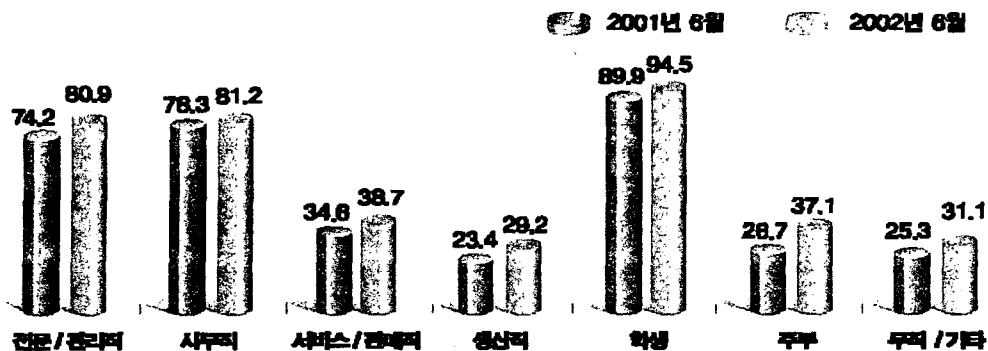
[표 1] 일반 1Mbps 이상 속도의 인터넷 서비스

구분	현황	세계수준	출처
인터넷 이용자(인구비율)	2,438만명(56.6%)	4~7위	'01.12 KRNIC
초고속인터넷* 가입자수	781만명	1	'02.2 MIC
인터넷 사용시간	19시간 20분	1	Nielsen/NetRatings
국가도메인수(ccTLD)수	45.7만개	5	'01.12 KRNIC

* 출처 : 정보화정책 제9권 제2호, 2002년 여름호

[직업별 인터넷 이용률]

(단위 : %)



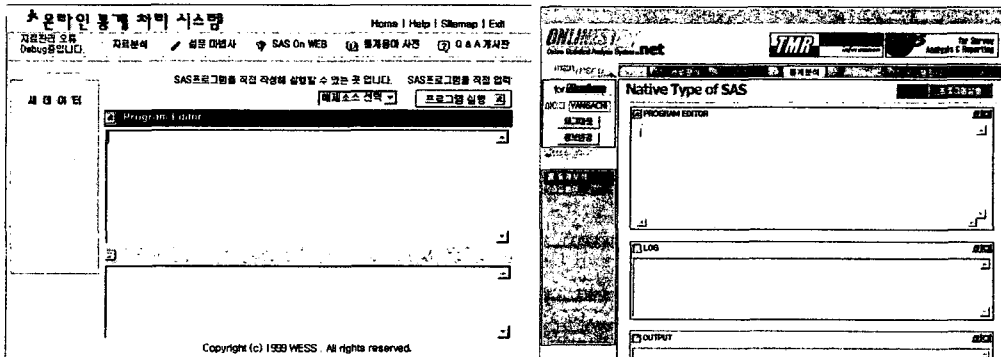
[그림1] 직업별 인터넷 이용률

* 출처 : 한국인터넷 정보센터 2002 한국인터넷 통계집

이러한 infra에 따라 웹상에서는 많은 통계관련 사이트들이 개설되어 있고 사용자들도 이를 통해 통계에 관한 정보를 얻고 있으며, 통계분석을 위한 소프트웨어들도 여러 사이트에서 WEB을 통한 서비스를 실행하고 있다. [그림2 참조]

그러나 이러한 통계분석 사이트들은 앞서 언급한 바처럼 프로그램을 직접 입력하여 분석하거나 데이터를 직접 업로드하여 분석하여야 하고 실행된 결과를 다시 데이터로 저장할 수 없는 그러한 단점을 지니고 있다. 따라서 대용량의 데이터 작업이나 보안을 요하는 데이터 분석, 혹은 분석결과를 받아서 추가적인 분석을 수행하고자 할 때에는 한계에 부딪히게 된다.

또한 프로그램을 직접 입력해야지만 실행할 수 있기 때문에 분석에 필요한 코드를 알고 있어야지만 정상적인 작업이 가능하다. 물론 차이는 있을 수 있지만 심도있는 분석이 아닌 기초통계학 수준의 통계분석을 위해서 코드를 배워야 한다는 것은 현 윈도우 GUI환경의 흐름에 반하는 것이라 할 수 있다.



1) 한남대학교 WESS

2) 대덕마케팅연구소 ONLINESTAT

[그림 2] Web에서 서비스되는 통계사이트

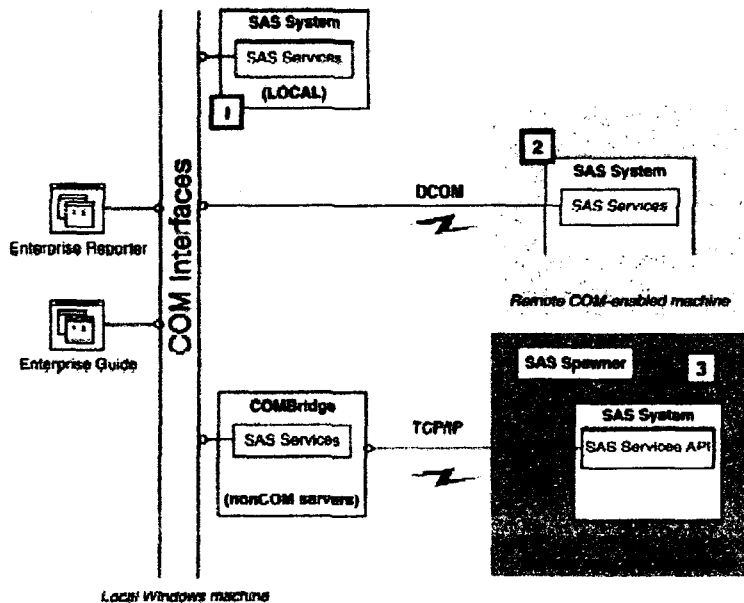
이러한 이유들로 인해 대부분의 통계분석은 통계 소프트웨어를 자신의 컴퓨터에 설치하고 분석을 수행하거나 마우스로 point & click할 수 있는 소프트웨어를 사용(예 : SAS, SPSS, Minitab, ...)하게 된다. 그러나 이와 같은 소프트웨어들을 사용하기 위해서 학교에서는 전산실에 직접 설치를 해주거나 학생은 개별적으로 이것을 구해서 자신의 집에 설치하고 있다. 이는 관리자의 입장에서는 소프트웨어의 유지 관리의 어려움을 초래하고 개인들은 필요할 때마다 학교에 요청하거나 개별적으로 불법적인 소프트웨어를 보유하게 만드는 결과를 가져온다. 또한 데이터의 양이 많아지면서 분석을 위한 데이터의 이동이 문제가 된다. 현재 이동이 편리한 매체(예 : 메모리стик)가 나와있기는 하지만 이를 개인이 일일이 복사해서 사용한다는 것은 무리가 있고 인터넷으로 다운로드받는다 할지라도 시간소요가 불가피하다.

이러한 문제점들로 인하여 얼마전 ASP(Application Service Provide)라는 개념의 소프트웨어의 관리, 분배 그리고 중앙데이터 센터를 통한 네트워크를 통해서 소비자들에게 소프트웨어를 제공하는 기술이 나왔으나 통계교육에 관련된 ASP가 없을 뿐만 아니라 기존 통계분석 소프트웨어에서 제공하는 심도있는 통계분석을 위한 요소까지 제공하기는 어렵다.

이러한 문제점들에 대해서 통계분석 소프트웨어 및 데이터를 중앙에서 집중관리하고 사용자들은 개인의 PC에는 기본적인 모듈만 설치하고 실제 분석은 분석서버에 접속하여 분석할 수 있는 형태의 Thin-Client 방식을 사용할 수 있다면 위에서 제기된 문제점들을 해결할 수 있다.

3. SAS를 새롭게 사용하는 방법 - Enterprise Guide

앞서 언급한 이러한 문제점들에 대해 새로운 대안으로 출시된 제품인 Enterprise Guide(이하 EG)는 세계적인 통계분석 소프트웨어 회사인 SAS에서 만든 Thin-Client 개념의 소프트웨어이다. Thin-Client란 Client(일반적으로 PC)에는 인터페이스만 설치하고 실제 필요한 모듈들은 서버에 탑재시켜 client와 server의 통신을 통한 작업을 수행하는 것을 의미한다. 이에 EG의 architecture는 [그림 3]와 같다.



[그림 3] EG Architecture

* 출처 : administrator for enterprise client : second edition(www.sas.com)

[그림3]에서 1은 일반적으로 자신의 PC에 SAS를 셋업하고 EG를 사용하는 것으로 이러한 방식은 현재 사용하고 있는 방식에 EG만 셋업을 하는 것이다. 이럴 경우 자신의 PC가 서버역할을 하는 형태가 되며 대부분의 통계 소프트웨어들은 이러한 방식을 사용하고 있다. 하지만 EG가 실제로 추구하는 방향은 2, 3과 같은 경우이다. 일단 이 방식들은 SAS가 서버에 설치되어 있다. 다시 말해 학교서버나 학과의 자체 서버에 SAS가 셋팅되어 있고 사용자는 자신의 PC에 EG만 설치한 다음 학교, 학과 서버에 접

속하여 서버의 자원을 활용하는 것이다. 이렇게 된다면 관리자는 서버만 제대로 관리를 해주고 사용자에게 대한 인증등록을 하고 데이터를 서버에 집중을 시킨다. 또한 사용자는 ftp와 같은 방식을 통하여 프로그램을 다운받아 자신의 PC에 설치하고 이를 서버와 연결만 시킨다면 서버의 막강한 자원(메모리, CPU 등)을 활용하여 분석할 수 있는 것이다.

3.1 EG의 장점

* 네트워크를 활용한 소프트웨어로서의 장점

한범수, 안정용, 한경수(2002)의 연구에서 제기된 통계소프트웨어들의 문제점 측면에서 바라보면 EG는 사용자가 setinit과 같은 라이선스를 갱신할 필요가 없어진다. 관리자가 서버에 있는 SAS의 라이선스만 갱신해 주면 사용자는 그 자원을 활용하는 것이므로 사용자가 일일이 신경쓸 필요가 없어지며 EG는 SAS의 기술력을 그대로 활용하는 것이므로 어떠한 데이터베이스라 할지라도 권한만 있다면 접근이 가능하다. 또한 사용자가 작업한 내용들이 프로젝트라는 개념으로 저장되어 지므로 자신의 분석 히스토리가 남게되며 이를 다른 이용자와 공유도 가능하다.

따라서 한범수, 안정용, 한경수(2002)의 연구에서 제시한 네트워크를 활용한 소프트웨어의 장점들을 활용할 수 있다.

* 중앙집중적인 관리

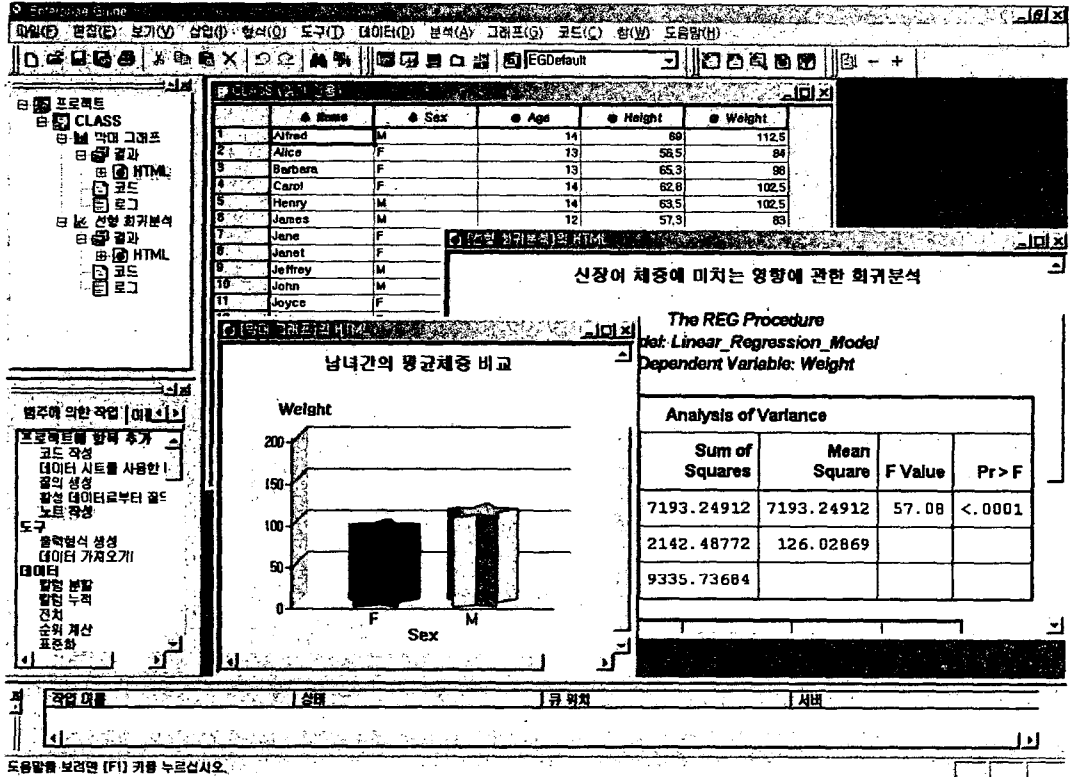
관리자는 그룹별로 사용자의 서버접근 권한을 줄 수 있다. 즉, 학과별 혹은 학년별 등의 그룹을 나누어 각 그룹별로 access할 수 있는 권한을 부여할 수 있으며, 사용자들을 각 해당그룹에 할당시킴으로 사용자가 EG를 실행하면 사용자의 권한에 알맞은 서버와 데이터를 보여준다. 관리자가 생성한 이런 정보들은 Repository라는 정보관리 데이터베이스에 저장되어지며, 사용자들은 관리자 소유의 이 repository에 원격으로 접속하여 자신의 작업환경을 가져올 수 있다.

* 사용의 편의성

화면 구성이 일반 사용자에게 친숙한 Graphics User Interface를 제공하며 마우스를 클릭 혹은 drag & drop 방식만으로도 쉽게 데이터 분석 및 그래프 생성, 레포트 출력 등이 가능하다.[그림 4 참조]

* 막강한 분석기능

EG는 SAS의 자원을 활용하는 것으로 SAS에서 분석되어 지는 모든 기법들(단변량 분석에서부터 다변량까지)을 사용할 수 있다. 물론 GUI환경의 소프트웨어들의 단점은 분석기능이 메뉴에서 한정되게 보이므로 이들을 보완하기 위해 추가적으로 코드를 삽입하여 분석을 할수있게 되어있다. EG도 이러한 기능이 제공되므로 Power-User들은 이 기능을 이용할 수 있으며 모든 작업들에 대한 source 코드가 생성되기 때문에 추가적으로 필요한 옵션 등을 직접 입력하여 더 심도깊은 결과도 얻을 수 있다.



[그림 4] Enterprise Guide 화면

* 깔끔한 그래프 및 다양한 레포팅 기능

EG는 기본적으로 ActiveX 기술을 이용하여 약 60여종의 그래프를 출력한다. 이 ActiveX는 웹브라우저가 셋업되어 있는 곳에서는 어디서나 사용할 수 있는 것으로 다이나믹한 그래프를 제공한다. 뿐만 아니라 자바 애플릿을 이용한 그래프, 전통적인 그래프 형태(JPEG, GIF)도 선택적으로 출력할 수 있다. 또한 기본적인 출력이 HTML 방식으로 웹브라우저만 있으면 이들 결과를 다이나믹한 형태 그대로 쉽게 볼 수 있으며, 사용자의 요구에 따라 PDF, Text, 기타 응용프로그램(예:엑셀)으로 레포팅을 할 수 있으며 이들 결과를 e-mail 등을 통하여 보낼수 있다.

* 확장성

만약 사용자가 필요한 작업을 GUI환경으로 만들고 싶다면 Visual Basic 과 같은 COM(Component Object Model)을 지원하는 개발도구를 사용해서 사용자의 요구에 맞는 작업들을 추가로 메뉴화 시킬 수 있다. 따라서 이를 사용하는 사용자나 관리자가 필요한 기능을 메뉴화시킴으로서 기존 GUI 환경의 소프트웨어에서 코드로 구현을 하거나 아예 분석을 포기한 작업들도 쉽게 GUI화 시킬 수 있다.

* 다양한 플랫폼

EG는 SAS기반의 애플리케이션이므로 SAS의 강점중 하나인 플랫폼에 영향을 받지 않는다. 다시 말해서, UNIX서버나 NT서버 어느것이나 사용할 수 있으며, 소규모의 과에서는 일반 PC급의 컴퓨터를 서버로 사용할 수도 있다.

3.2 문제점

* 네트워크 애플리케이션이다.

이것은 반드시 네트워크로 연결되어 있어야 한다는 것이다. 하지만 이것은 현재 사용되고 있는 무선랜 기술이 더욱 발전하면 자연스럽게 해결되어질수있다. 또한 LAN으로만 연결된 곳에서도 충분히 EG의 Architecture를 사용하여 구축할 수 있다.

* OS의 보안체계를 따른다.

EG는 자체적으로 보안관리를 하지 않는다. 앞서 언급한 바와 같이 그룹, 사용자 분류만 가능하기 때문에 서버로의 접속권한은 그 서버의 OS 체계에 따르게 되어 있다. 이는 관리자가 서버상에서 사용자에게 권한을 직접 설정해 주어야 한다는 것이다. 하지만 이 역시 서버관리의 기본 작업이므로 오히려 이것은 별도의 추가적인 권한 할당 작업이 필요없음을 의미하며 관리자는 단지 이미 주어진 사용자들의 권한에 맞게 그룹핑만 해 주면 되는 것이다.

* 네트워크에 많은 부하를 줄 수 있다.

이것은 어떤 네트워크 애플리케이션이라 할지라도 풀어야 하는 숙제 중 하나이다. 사용자가 늘어나고 동시접속이 늘어난다면 이것은 어쩔 수 없는 문제일지도 모르지만 일반 학교 웹사이트의 접속이 아닌 통계분석을 위해 동시접속자가 수백명 이상이 된다는 것은 현실적으로 일어나지 않으리라 생각한다.

4. 결론

위에서 통계분석 소프트웨어로서의 EG의 모든 기능을 설명할 수 없었지만 EG의 사상 및 통계교육에서의 활용방안 등에 대해 정리해 보았다. 앞으로도 인터넷 환경은 더욱 좋아질 것이며 대용량 데이터를 이용한 통계분석에 대한 사용자의 욕구도 점차 증가할 것이므로 이를 활용하여 학습자가 쉽고 편리하게 통계분석을 할 수 있을 것이다.

참고문헌

1. Beom Soo Han, Jeong Yong Ahn and Kyung Soo Han (2002). On the Design of Statistical Software in the Network Environment. *The Korean Communications in Statistics*, Vol. 9, No 1, 167-174
2. Carol Rigsbee, Chris Hemedinger (2001). "Delivering Information to the People Who Need to Know", SUGI 26 Conference proceedings.
3. 한국인터넷 정보센터 2002 한국인터넷 통계집
4. 정보화정책 제9권 제2호, 2002년 여름호