

Development of Multivariate Analysis System by Using SAS/AF and SCL¹⁾

Sang-Tae Han²⁾, Hyuncheol Kang³⁾, Seong-Keon Lee⁴⁾
Myung-Seok Jang, Duck-Ki Lee, Dong-Kyun Ryu⁵⁾

Abstract

In recent years, the development and the embodiment of information analysis system has been sprightly carried out in several fields of study. In this study, as an extension of these studies, we develop a system for multivariate analysis which might be widely used in social and natural sciences. This multivariate analysis system is developed by using multivariate analysis procedures in SAS/STAT software. Also, the system supply users with the environment of GUI(Graphical User Interface), which is constructed with AF(application frame) and SCL(screen control language) of SAS software, in order to help users to use the system with easy.

Keywords : Multivariate Analysis System, Application Frame, Screen Control Language

1. 서론

최근에 다변량분석과 관련하여 시스템을 구현하는 연구가 다양한 각도로 이루어졌다. 서혜선 · 김미경 · 허명희(1999)는 허명희(1999)에 의해 연구된 다변량 수량화 분석을 시스템으로 구현하였으며, 현기홍 · 최용석(2000)은 이원표 자료행렬의 행과 열을 그래프에 동시에 나타내어 이들의 관계와 패턴을 알아보기 위한 다변량 그래프적 분석기법인 행렬도(biplot)를 시스템으로 구현하였다. 이들이 구현한 시스템은 기존의 전통적인 다변량분석의 논리와 차별화 된 자료 탐색적 관점의 논리를 시스템으로 구현한 특징을 가지고 있다. 이러한 시스템 개발의 연장선상에서 SAS를 이용하여 다변량 자료분석을 수행할 때 사용자들이 겪는 실제적 어려움을 다소나마 해결해 주고자 하는 것이 본 연구의 주요목적이라 할 수 있다.

통계적 자료분석을 위해 가장 널리 사용되고 있는 SAS는 다양한 분석을 위한 처리기능을 가지

1) 이 논문은 2000년도 호서대학교 교내연구비 지원에 의하여 연구되었음

2) Assistant Professor, Department of Mathematics, Hoseo University, Asan 336-795, Korea.
E-mail : sthan@office.hoseo.ac.kr

3) Lecture Professor, Department of Mathematics, Hoseo University, Asan 336-795, Korea.

4) Department of Statistics, Korea University, Seoul 136-701, Korea.

5) Department of Mathematics, Hoseo University, Asan 336-795, Korea.

고 있으나, 사용법을 습득하고 출력결과를 해석하기 위해 적지 않은 노력이 필요한 것이 사실이다. 따라서 많은 사용자들이 SAS를 이용하여 다변량분석을 수행할 때 갖는 가장 큰 어려움은 SAS/STAT의 다변량분석 프로시저들에 대한 다양한 지식과 원하는 그래프를 예쁘게 출력시킬 수 있는 SAS/GRAFH 등에 대한 상당한 수준의 프로그래밍 능력이 요구된다는 것이다. 실제적으로 다변량 통계방법론에 대한 상당한 수준의 통계학 실력을 갖추는 것도 쉽지 않은 일반 사용자 입장에서 SAS 프로그래밍에 대한 능력을 추가적으로 요구받는 것은 커다란 부담이 아닐 수 없다.

이런 관점에서 SAS 등 통계 소프트웨어에 익숙하지 않은 일반 사용자들도 다변량분석을 쉽게 활용할 수 있도록 메뉴방식의 시스템을 개발하여 보급하고자 하는 것이 본 연구의 주요 목적이라 할 수 있으며, 본 시스템을 이용함으로써 그 동안 SAS의 메뉴방식에서 제공해 주지 못해(주성분 분석 등 일부분석만 제공되고 있음) 충분한 활용이 부족했던 여러 다변량 분석방법들을 보다 쉽게 활용할 수 있을 것이다.

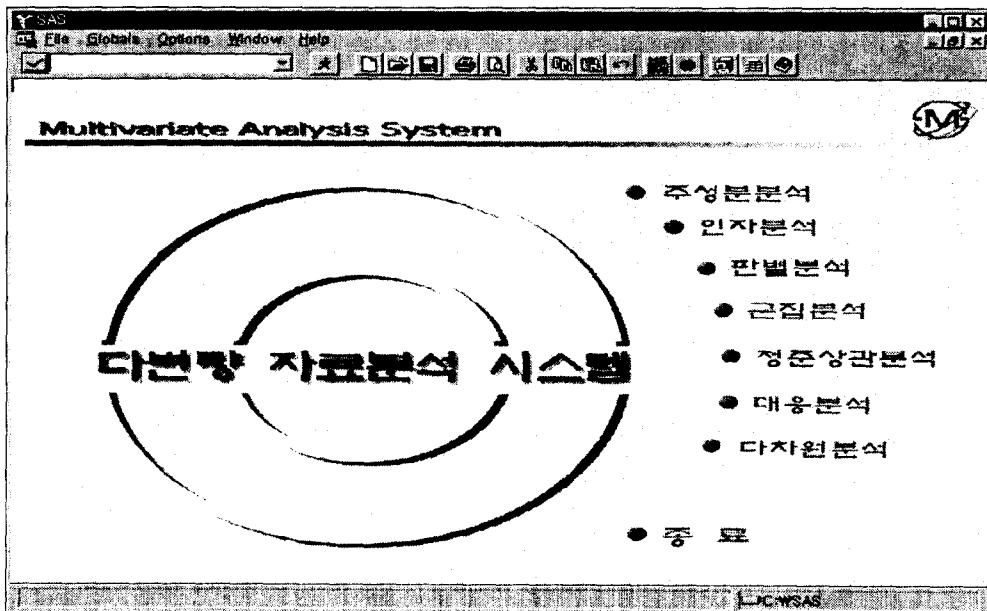
2. 다변량 자료분석 시스템의 구성

다변량 자료분석 시스템은 그래픽 사용자 인터페이스(Graphical User Interface) 하에서 SAS 소프트웨어에서 제공되는 여러 가지 기능을 이용하여 특별한 목적에 맞게 응용프로그램을 개발할 수 있도록 지원하는 객체지향 도구인 SAS/AF(application frame)와 이를 화면의 뒤에서 응용프로그램이 수행될 수 있도록 도와주는 화면제어 언어인 SCL(screen control language)을 이용하여 개발되었다(강현철·한상태 외, 1999; SAS Institute, 1993, 1994).

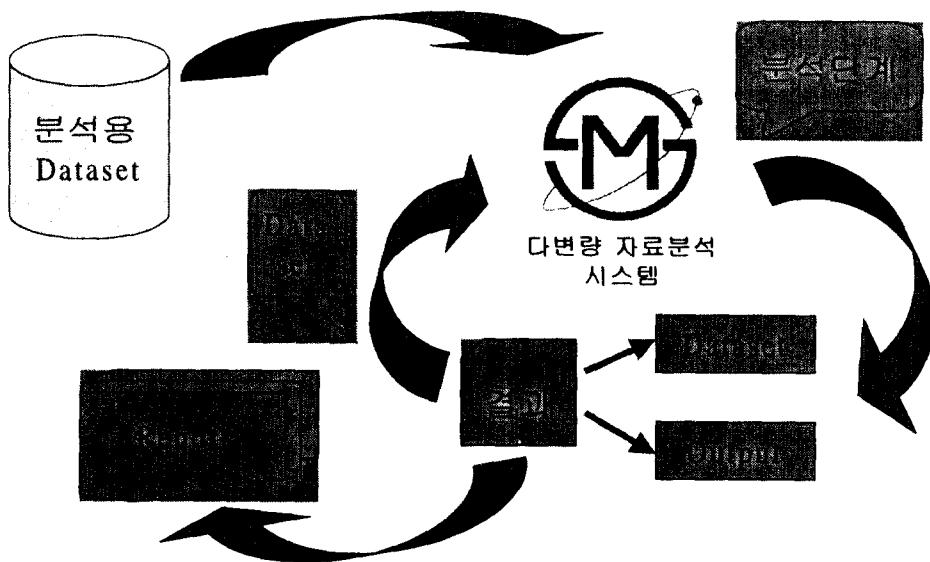
본 시스템은 SAS/STAT, GRAPH, IML을 기초로 SAS/AF 및 SAS/SCL을 활용하여 개발되었으며 크게 다음과 같은 세 가지 부분으로 구성되어 있다. 첫 번째 부분은 응용프로그램의 흐름에 맞게 생성한 화면인 프레임으로 사전에 정의된 또는 사용자가 생성한 객체(object)들로 이루어져 있다. 두 번째 부분은 프레임과 프레임의 연결, 각 객체의 작동방법(method) 등에 대한 프로그램을 저장해 놓은 것으로 화면과 통계분석의 수행(submit)을 연동시켜주는 부분인 SCL 코드이다. 마지막은 SCL 내에 존재하며 실질적 통계분석이 수행(submit)되는 부분으로 SAS/STAT, SAS/GRAFH, SAS/IML 등으로 이루어져 있다.

본 연구에서 개발한 다변량 자료분석 시스템은 다변량 자료분석의 대표적인 방법들인 주성분분석(Principal Component Analysis), 인자분석(Factor Analysis), 판별분석(Discriminant Analysis), 군집분석(Cluster Analysis), 정준상관분석(Canonical Correlation Analysis), 대응분석(Correspondence Analysis), 다차원 척도법(Multidimensional Scaling) 등을 분석할 수 있도록 되어 있다. <그림 2.1>은 다변량 자료분석 시스템의 초기화면이다. 또한 각 분석방법은 <그림 2.2>에서 볼 수 있는 바와 같이 분석용 데이터세트(data set)를 본 시스템에 적용하고 미리 짜여진 분석단계에 의해 결과물을 산출하고, 그 결과물을 보고서 형태로 출력할 수 있으며, 다시 본 시스템에 적용하여 새로운 분석방법을 적용시킬 수 있다.

<그림 2.1> 다변량 자료분석 시스템의 초기화면



<그림 2.2> 다변량 자료분석 시스템의 구성



3. 다변량 자료분석 시스템 사용법

3.1 시작화면

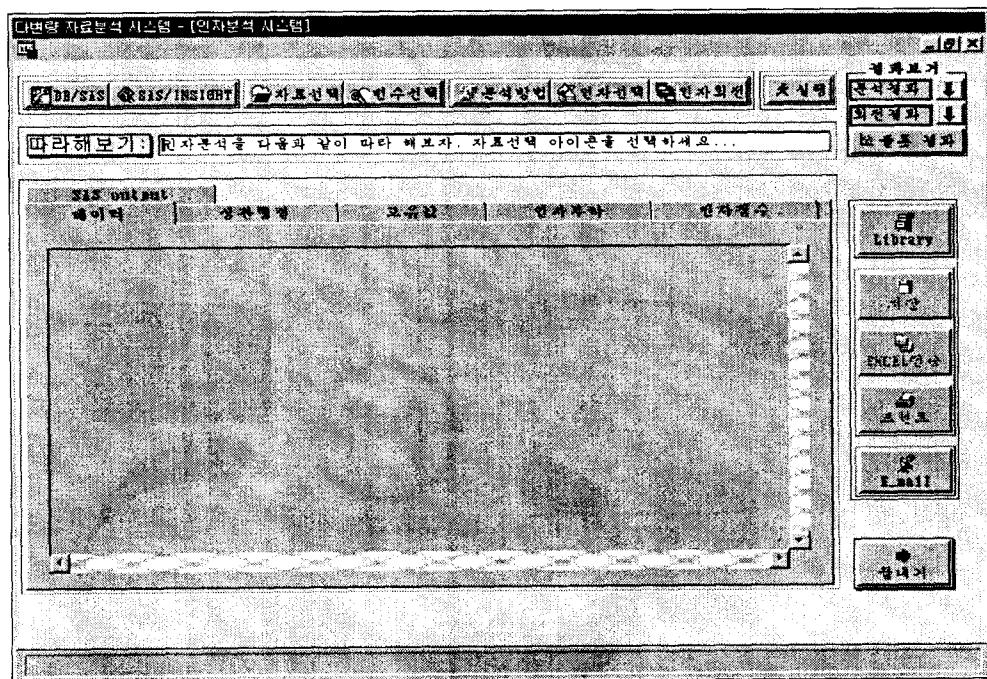
다변량 자료분석 시스템에 들어가는 방법은 두 가지가 있다. 첫 째는 윈도우의 바탕화면에 '시

스템의 아이콘을 만들어 놓은 후 “바로가기 아이콘”을 클릭하는 방법이고, 두 번째 방법은 SAS를 구동시킨 후에 라이브러리를 지정한 후 SAS의 명령어 상자(command box)나 명령어 줄(command line)에서 ‘af c=mult.anly.start,frame’을 입력한 후 [Enter] 키를 누르는 방법이다. 어느 방법을 이용하더라도 시스템의 초기화면인 <그림 2.1>이 나타난다.

3.2 분석화면과 종료화면

초기 화면에서 사용자가 원하는 분석방법을 클릭하면 해당되는 다변량 기법 분석화면이 나타나게 된다. <그림 2.1>에서 인자분석을 클릭하여 나타난 인자분석 시스템이 <그림 3.1>이다. 또한 초기화면에서 종료를 선택하면 다변량 자료분석 시스템을 빠져 나가게 된다.

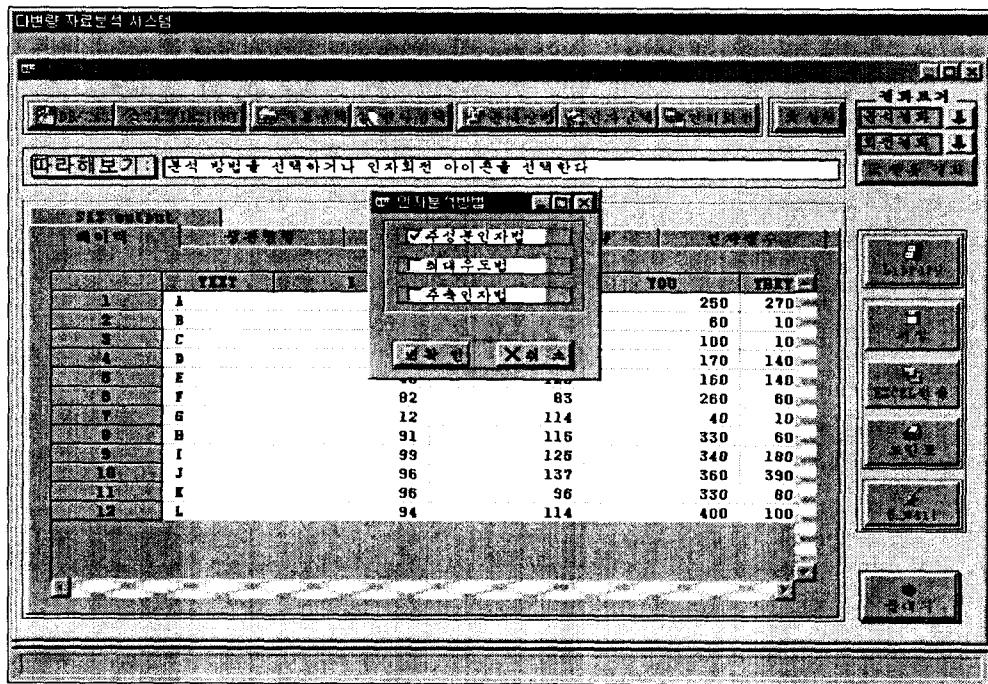
<그림 3.1> 인자분석 시스템 화면



3.3 분석화면

사용자가 지정한 분석화면이 등장하면 사용자는 분석에 필요한 각각의 선택버튼을 눌러 원하는 옵션을 지정함으로서 분석을 마칠 수 있다. 예를 들어, 인자분석 시스템 화면에서 순서대로 자료선택 버튼을 클릭하여 분석할 자료를 지정하고, 분석변수 버튼을 클릭하여 변수를 지정하며, 분석방법 버튼을 클릭하여 인자의 추정방법을 선택하고(주성분인자법, 최대우도법, 주축인자법 중 복수선택 가능), 실행 버튼을 클릭하여 분석을 수행할 수 있다. 또한 인자분석에서는 인자회전 버튼을 클릭하여 회전방법을 선택하고(복수선택 가능), 실행 버튼을 클릭하면 회전된 분석결과를 얻을 수도 있다.

<그림 3.2> 인자의 추정방법 선택 화면



3.4 결과보기

결과보기는 <그림 3.2>의 우측상단에 결과보기 메뉴에서 분석결과를 선택하면 인자분석 결과화면에 테이블 형태로 분석결과(상관행렬, 고유값, 인자부하, 인자점수 등)가 정리되어 나타나게 된다. 또한 추가적으로 SAS output 탭을 누르면 SAS output 윈도우에 출력되는 형태대로 결과를 얻을 수도 있다. 만약 인자회전된 분석결과를 보기 위해서는 결과보기에서 회전결과를 선택하면 된다. 이들 분석결과는 SAS 데이터 파일이나 텍스트 파일로 저장할 수 있다.

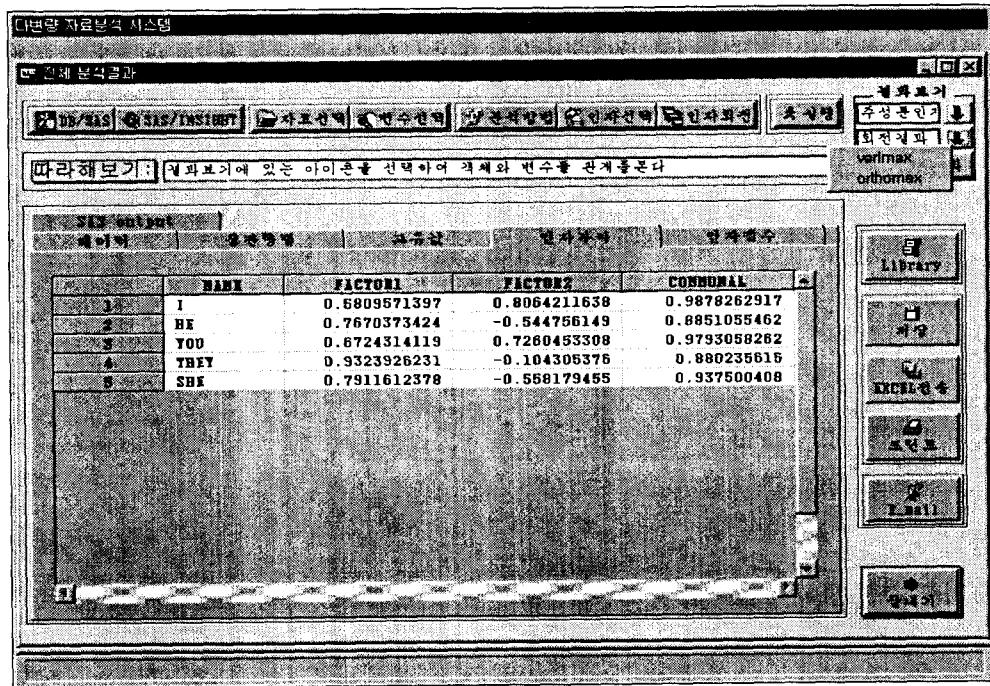
또한 결과보기에서 플롯 그리기를 선택하면 <그림 3.4>에서와 같이 인자분석 결과에 대한 2차원 그래프 또는 3차원 그래프를 볼 수 있다. 여기서는 각 추정방법 및 회전방법에 대하여 개체플롯과 변수플롯을 통해 결과를 시각적으로 살펴 볼 수 있을 뿐만 아니라, 회전 전후의 결과를 비교하거나 그래프를 확대·축소하는 것(<그림 3.4>의 실행메뉴 우측 버튼 이용)도 가능하다.

3.5 기타

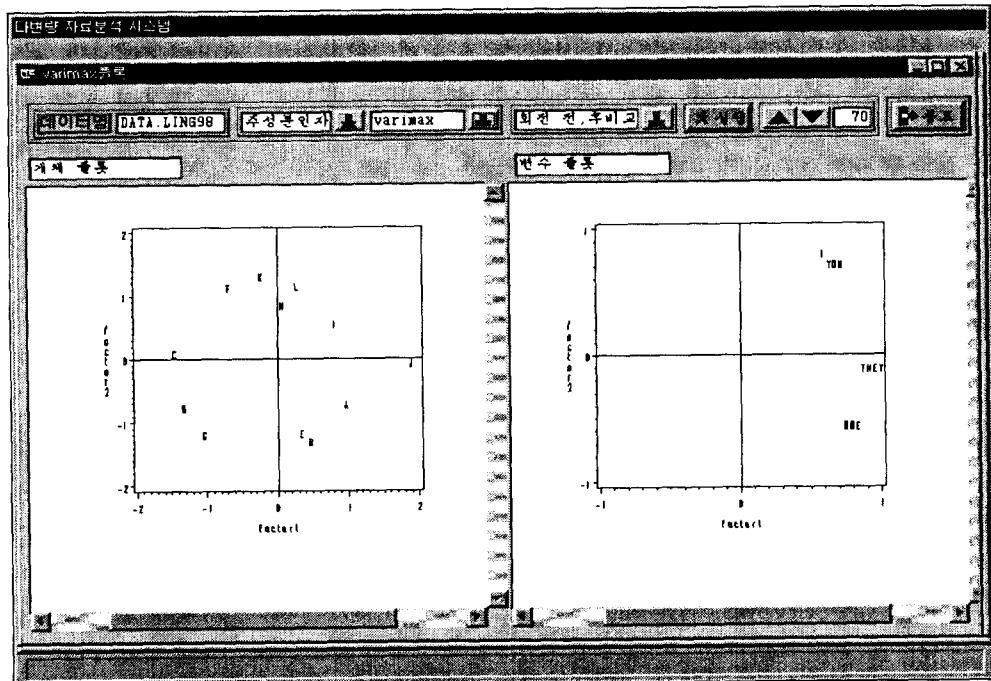
다면량 자료분석 시스템에는 이외에도 데이터 및 결과의 입출력을 담당하는 여러 가지의 부수적인 기능들이 포함되어 있다. 예를 들어, DB/SAS 버튼은 SAS/Access와 연동되어 있어 외부데이터베이스 파일을 SAS 데이터 파일로 변형하는 기능을 가지고 있고, SAS/Insight 버튼은 SAS/Insight와 연동되어 다양한 그래프 그리기 및 기초통계분석을 수행할 수 있는 기능을 가지고 있으며, 저장 버튼은 분석결과를 텍스트 파일 또는 SAS 데이터세트로 저장하는 기능을 담당하고, EXCEL 전송 버튼은 MS Office의 엑셀과 연동되어 분석결과를 엑셀로 전송하는 기능을 담당하고 있다. 또한 E-mail 버튼을 클릭하여 다음 <그림 3.6>에서와 같이 분석결과를 전자메일로 바로

전송할 수도 있다.

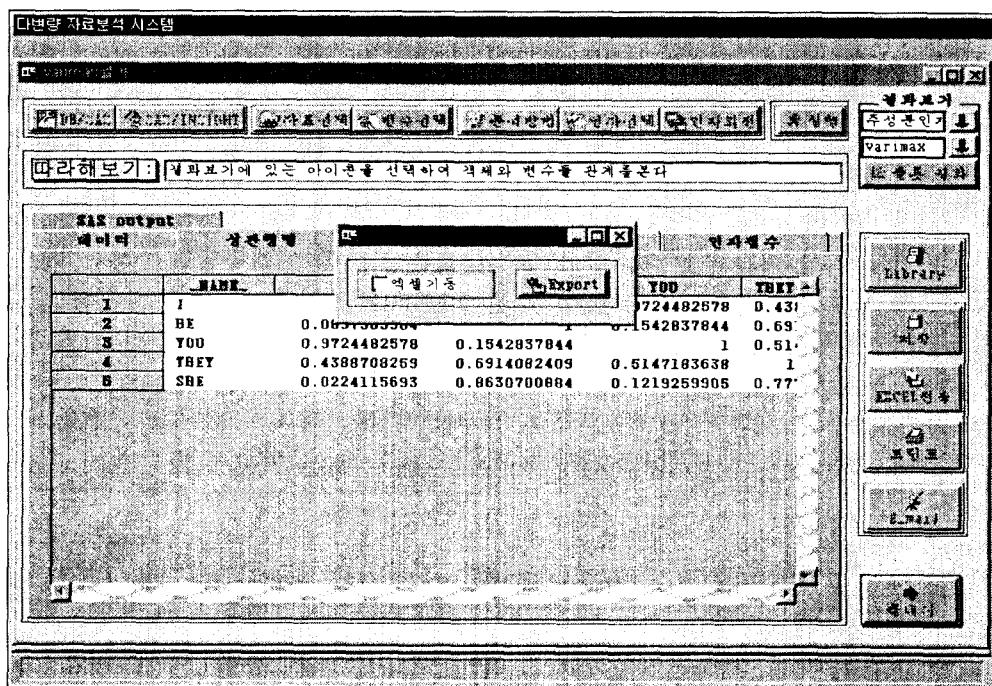
<그림 3.3> 인자분석의 결과보기 화면



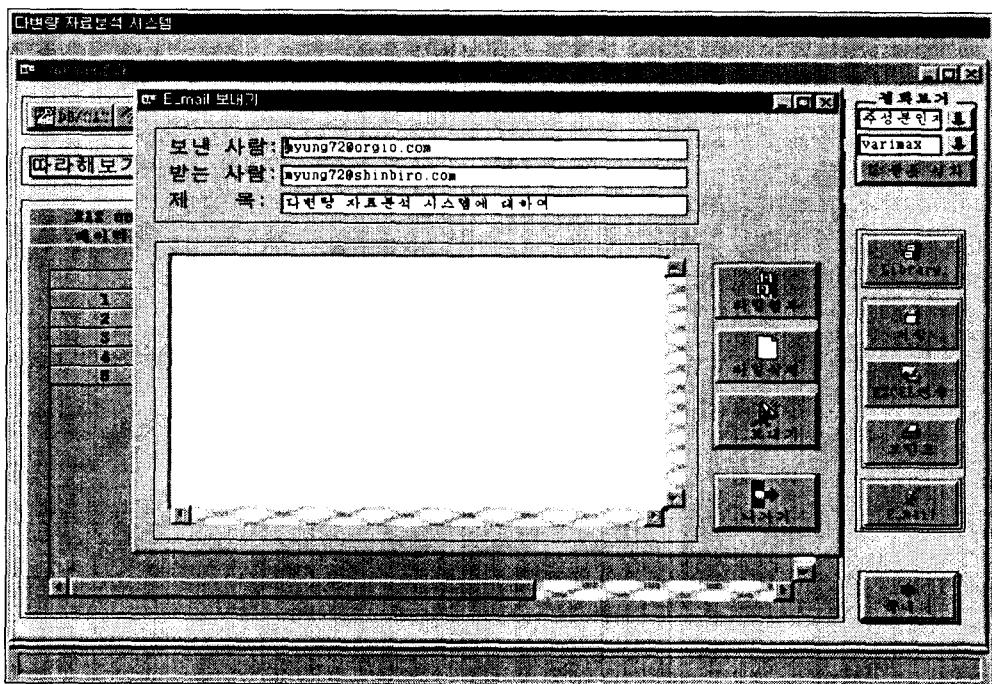
<그림 3.4> 그래프 보기 화면



<그림 3.5> 엑셀로 결과 보내기



<그림 3.6> E-mail 보내기



4. 결론

본 연구에서는 사회과학 및 자연과학에서 널리 사용되는 7가지 다변량 분석방법을 일반 사용자가 쉽게 적용할 수 있도록 메뉴방식의 시스템을 구축하였다. 앞 절에서는 인자분석 시스템의 예를 들어 본 다변량 자료분석 시스템의 기능을 간단히 설명하였으며, 다른 분석방법에 대해서도 이와 유사한 과정을 통해 분석방법을 적용하고 결과를 얻는 과정을 쉽게 수행할 수 있다. 대부분의 일반 사용자는 SAS의 언어 및 방대한 프로시저의 사용을 어려워하는 것이 현실인데, 본 다변량 자료분석 시스템이 그러한 어려움을 다소나마 덜 수 있을 것이라고 생각된다. 다만 앞으로 본 시스템을 계속적으로 사용하면서 발견되는 문제점 및 사용상의 불편한 점 등을 보완하고 일반 사용자들의 편의를 위해 풍부한 도움말을 시스템에 포함시켜, 본 시스템이 보다 편리한 환경에서 쉽게 사용될 수 있도록 지속적으로 개선해 나갈 생각이다.

마지막으로 본 연구에서 개발한 다변량 자료분석 시스템을 사용하기 원하는 사용자는 저자들의 e-mail(sthan@office.hoseo.ac.kr, hychkang@office.hoseo.ac.kr)로 문의하기 바란다.

참 고 문 현

- [1] 강현철 · 한상태 · 서혜선 · 정형철 · 최보승 (1999). 「통계적 정보분석시스템 개발」, 자유아카데미, 서울.
- [2] 서혜선 · 김미경 · 허명희 (1999). SAS AF/SCL로 구현한 다변량 수량화 시스템, 「한국분류학회지」, 제3권, 1-11.
- [3] 허명희 (1999). 「다변량수량화」, 자유아카데미, 서울.
- [4] 현기홍 · 최용석 (2000). 행렬도 시스템(Biplots System)의 개발, 「응용통계연구」, 제13권, 297-306.
- [5] SAS Institute Inc. (1993). *SAS/AF Software: FRAME Entry Usage and Reference, Version 6, First Edition*, SAS Institute Inc., NC: Cary.
- [6] SAS Institute Inc. (1994). *SAS Screen Control Language: Reference, Version 6, Second Edition*, SAS Institute Inc., NC: Cary.