

## EXCEL을 이용한 다변량자료분석 시스템 개발

한상태<sup>1)</sup> 강현철<sup>2)</sup> 한정훈<sup>3)</sup>

### 요약

최근 다변량자료분석과 관련하여 이를 시스템으로 구현하려는 연구가 다양한 각도로 이루어지고 있다. 이러한 연구들의 공통적인 특징은 일반 사용자들에게 고급 통계분석기법을 편리하게 활용할 수 있도록 GUI(Graphical User Interface) 환경의 시스템을 제공해 준 것이다. 이러한 연구들의 연장선 상에서, 본 연구에서는 사회 각 분야에서 가장 널리 활용되고 있는 사무용 프로그램인 Excel을 활용하여 시스템을 개발함으로써, 일반 사용자들도 대화식으로 다변량자료분석을 쉽게 수행할 수 있도록 하였다.

주요용어: 다변량자료분석 시스템, Excel Macro, VBA

### 1. 서론

다변량자료분석과 관련하여 시스템을 구현하는 연구가 다양한 각도로 이루어졌다. 서혜선 등(1999)은 허명희(1999)에 의해 연구된 다변량 수량화분석을 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 구현하였고, 한상태 등(2001)은 SAS AF와 SCL을 이용하여 SAS의 메뉴방식에서 제공해 주지 못해 충분한 활용이 어려웠던 다변량기법들을 일반 사용자들도 쉽게 사용할 수 있는 메뉴방식의 다변량자료분석 시스템을 개발하였다.

통계자료분석을 위해 가장 널리 사용되고 있는 SAS 및 SPSS와 같은 통계패키지는 다양한 분석을 위한 처리기능을 가지고 있으나 이에 대한 사용법을 습득하고 출력결과를 해석하기 위해서는 많은 시간과 노력을 기울여야하고, 또한 소규모 기업이나 개인이 구입해서 사용하기에는 가격이 비싸서 일반 사용자에게는 커다란 부담이 되고 있다. 이런 문제들을 해결하기 위해 조신섭 등(1998)은 Excel을 기반으로 하여 기초통계분석을 처리할 수 있는 'KORSTAT' 시스템을 개발하였다. 그러나 이 시스템은 현재 Excel 도구에 있는 '데이터 분석' 기능에서 처리할 수 있는 기초통계분석이 대부분이라 할 수 있다. 즉 Excel을 기반으로 하는 다변량자료분석 기능은 갖추고 있지 못한 문제점을 안고 있다.

이와 같은 제한점을 극복하기 위해 본 연구에서는 사회 각 분야에서 가장 널리 활용되고 있는 사무용 프로그램 중 하나인 마이크로소프트(Microsoft) Excel을 활용하여 다변량

1) (336-795) 충청남도 아산시 배방면 세출리, 호서대학교 자연과학부 정보통계학전공, 부교수

E-mail: sthan@office.hoseo.ac.kr

2) (336-795) 충청남도 아산시 배방면 세출리, 호서대학교 자연과학부 정보통계학전공, 전임강사

E-mail: hychkang@office.hoseo.ac.kr

3) (336-795) 충청남도 아산시 배방면 세출리, 호서대학교 대학원 통계전공, 석사졸업

E-mail: cooljack94@empal.com

자료분석 시스템을 개발하고자 한다. Excel은 통합문서 프로그램으로 편리한 사용자 환경과 계산, 그래프 작성 기능 및 다양한 통계함수를 가지고 있다. 또한 기본적으로 기초통계, t-검정, 회귀분석, 분산분석 등을 ‘데이터 분석도구’라는 추가기능에서 제공하고 있으며, 이를 이용하면 대화식으로 통계분석을 수행할 수 있는 장점을 가지고 있다. 그러나, Excel에서 제공하고 있는 ‘데이터 분석도구’에서도 다변량자료분석을 수행할 수 있는 기능을 갖고 있지 못한 제한점이 있다.

이런 관점에서 본 연구에서는 Excel의 매크로(Macro) 기능과 VBA(Visual Basic for Application)를 이용하여 ‘데이터 분석도구’와 같은 추가기능 파일 형태(.xla)로 다변량자료분석 시스템을 개발하였으며, 설치방법과 사용법 또한 사용자의 편리를 위해 Excel의 ‘데이터 분석도구’ 기능과 동일하게 구성하였다. 따라서 개발된 시스템을 이용하면 일반사용자들이 굳이 통계패키지의 사용법을 익히지 않고서도 쉽게 다변량자료분석을 수행할 수 있는 장점을 제공받을 수 있을 것이다.

## 2. 다변량자료분석 시스템의 구성

다변량자료분석 시스템은 Excel의 매크로 기능과 VBA(Codemia, 2002)를 이용하여 Excel의 추가기능 파일 형태로 개발되었다. VBA는 객체지향 프로그램 언어인 Visual Basic을 기반으로 한 것으로써, 본질적으로는 Visual Basic과 동일하지만 Windows에서 단독으로 작동하는 프로그램을 제작할 수는 없고, Excel, Access, Word, Powerpoint, Outlook 등의 호스트 응용프로그램을 위한 프로그래밍 언어라 할 수 있다.

Q39	A	B	C	D	E	F
	City	Bread	Burger	Milk	Orange	Tomato
2	ATLA	24.5	94.5	73.9	80.1	41.6
3	BALT	26.5	91	67.5	74.6	53.3
4	BOST	29.7	100.8	61.4	104	59.6
5	BUFF	22.8	86.6	65.3	118.4	51.2
6	CHIC	26.7	86.7	62.7	105.9	51.2
7	CINC	25.3	102.5	63.3	99.3	45.6
8	CLEV	22.8	88.8	52.4	110.9	46.8
9	DALL	23.3	85.5	62.5	117.9	41.8
10	DETR	24.1	93.7	51.5	109.7	52.4
11	HONA	29.3	105.9	80.2	133.2	61.7
12	HOUS	22.3	83.6	67.8	108.6	42.4
13	KANS	26.1	89.9	65.4	100.9	43.2
14	LOSA	26.9	89.3	56.2	82.7	38.4
15	MILW	20.3	89.6	53.8	111.8	53.9
16	MINN	24.6	92.2	51.9	106	50.7
17	NEWY	30.8	110.7	66	107.3	62.6
18	PHIL	24.5	92.3	66.7	98	61.7
19	PITT	26.2	95.4	60.2	117.1	49.3
20	STLO	26.5	92.4	60.8	115.1	46.2
21	SAND	25.5	83.7	57	92.8	35.4
22	SANF	26.3	87.1	58.3	101.8	41.5
23	SEAT	22.5	77.7	62	91.1	44.9

그림 2.1: Excel에 설치된 다변량자료분석 시스템

표 2.1: 다변량자료분석 시스템의 분석방법 및 출력결과

주성분분석	인자분석	대응분석	정준상관분석	판별분석	다차원척도법
기초통계량	기초통계량	카이제곱통계량	정칙치, 고유치	정칙치, 고유치	고유치
공분산행렬	공분산행렬	행 가중평균	정준계수	정준계수	다차원 좌표
상관행렬	상관행렬	행 프로파일	행좌표, 열좌표	행좌표, 열좌표	스트레스
고유치	고유치	정칙치, 고유치	행렬도	행렬도	행렬도
고유벡터	인자적재	행좌표, 열좌표			
주성분점수	인자점수	행렬도			
행렬도	행렬도				

본 시스템은 크게 두 부분으로 구성되어 있다. 첫 번째 부분은 응용프로그램의 흐름에 맞게 생성한 화면인 사용자 정의 폼(form)이다. 사용자 정의 폼은 사용자가 생성한 객체(object)들을 자유롭게 정의하여 Excel의 워크시트와는 별도로 표시할 수 있는 독자적인 인터페이스를 의미하고, 각 객체 내에서 Visual Basic 코드로 기술되어 있는 처리를 실행하는 부분을 말한다. 두 번째 부분은 시스템을 기동시키기 위한 매크로 구문과 Excel 기능에는 없는 고유값-고유벡터를 찾기 위한 함수 및 그래프에 레이블을 붙이기 위한 함수가 정의되어 있는 모듈을 말한다.

본 연구에서 개발한 다변량자료분석 시스템은 주성분분석, 인자분석, 대응분석, 정준상관분석, 정준판별분석, 다차원척도법 등 다양한 분석을 수행할 수 있도록 되어 있다(그림 2.1은 다변량자료분석 시스템이 Excel에 설치되어 있는 화면이고, 표 2.1은 각 분석방법에서 출력할 수 있는 결과들이다). 또한 각 분석방법은 Excel의 워크시트에 입력된 데이터를 마우스로 드래그하거나 입력범위를 직접 입력함으로써 간단하게 시스템에 적용시킬 수 있고, 정해진 분석단계에 의해 결과물과 그래프를 새로운 워크시트에 출력할 수 있도록 되어 있다.

### 3. 다변량자료분석 시스템의 설치 및 실행

이 절에서는 개발된 다변량자료분석 시스템을 윈도우즈에 설치하는 과정을 설명하고, 주성분분석과 대응분석의 예를 들어 다변량자료분석 시스템을 사용하는 과정을 소개할 것이다. 인자분석, 정준상관분석, 판별분석, 다차원척도법 등도 유사한 과정을 통해 쉽게 분석방법을 적용하고 결과를 얻을 수 있다.

#### 3.1. 시스템 설치 및 제거 방법

시스템의 설치방법은 윈도우즈의 버전에 따라 다소 차이가 있는데, 그 이유는 MS Office 프로그램의 Addin 폴더의 위치 때문이다. 윈도우즈98에서는 'C:\Windows\Application Data\Microsoft\AddIns' 폴더에 개발된 '다변량자료분석시스템.xla'를 복사하여 붙여넣기를 한 후, Excel의 '도구→추가기능→찾아보기'에서 '다변량자료분석시스템.xla'를 선택하

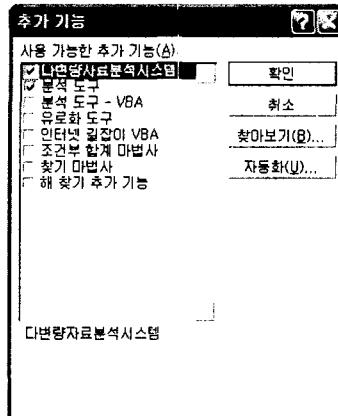


그림 3.1: 추가기능 대화상자에 다변량자료분석 시스템이 추가된 모습

고 확인을 클릭하면 그림 3.1과 같이 Excel의 추가기능 대화상자에 다변량자료분석 시스템 추가기능이 설치되는데, 여기서 확인 버튼을 클릭하면 앞의 그림 2.1과 같이 Excel의 풀-다운 메뉴에 다변량자료분석이라는 메뉴가 생성되게 된다. 또한 윈도우즈2000과 윈도우즈Xp에서는 'C:\Documents and Settings\사용자 폴더\Application Data\Microsoft\AddIns' 폴더에 '다변량자료분석시스템.xla'파일을 복사한 후 윈도우즈98에서와 같은 방법으로 설치를 하면 된다. 반대로 설치된 추가기능을 해제할 때는 '도구→추가기능'에서 '다변량자료분석 시스템' 체크박스의 체크(✓)된 부분을 해제시키면 된다.

### 3.2. 주성분분석의 수행 예

본 시스템에서 주성분분석을 수행하기 위해서는 Excel을 기동 한 후 나타나는 그림 2.1의 풀-다운 메뉴에서 다변량자료분석 메뉴를 선택한 후 주성분분석을 클릭하면 된다. 그림 3.2는 그림 2.1에서 주성분분석을 클릭하여 나타난 '주성분분석' 대화상자를 보여주고 있다.

주성분분석을 수행하기 위해서는 이 대화상자에서 자료의 입력범위를 지정하고 자료의 형태에 맞게 옵션을 설정하면 된다. 예를 들어, 자료의 첫째 행과 열에 변수이름과 개체이름이 포함되어 있다면 '첫째 행 변수이름 사용'과 '첫째 열 개체이름 사용' 옵션을 선택하여 준다. 또한 개체플롯과 변수플롯의 출력을 원한다면 출력옵션에서 '변수, 개체 플롯의 출력' 옵션을 선택하고 확인을 클릭하면 주성분분석의 결과, 변수플롯, 개체플롯이 그림 3.3과 같이 새로운 워크시트에 출력된다. 여기서 분석에 사용된 예제 데이터는 미국 23개 도시의 5개 품목에 대한 물가자료이다(허명희, 1999). 주성분분석 대화상자를 통해 얻을 수 있는 출력결과란 각 변수의 평균, 공분산행렬, 상관행렬, 고유값, 누적점유율, 고유벡터, 주성분 점수, 행렬도 등이다.

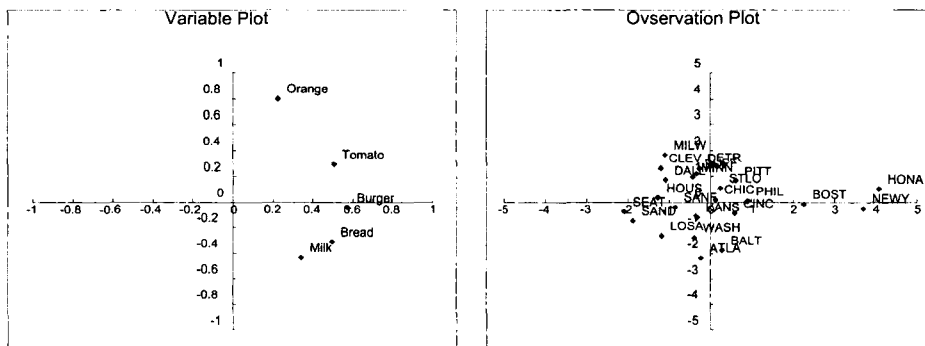
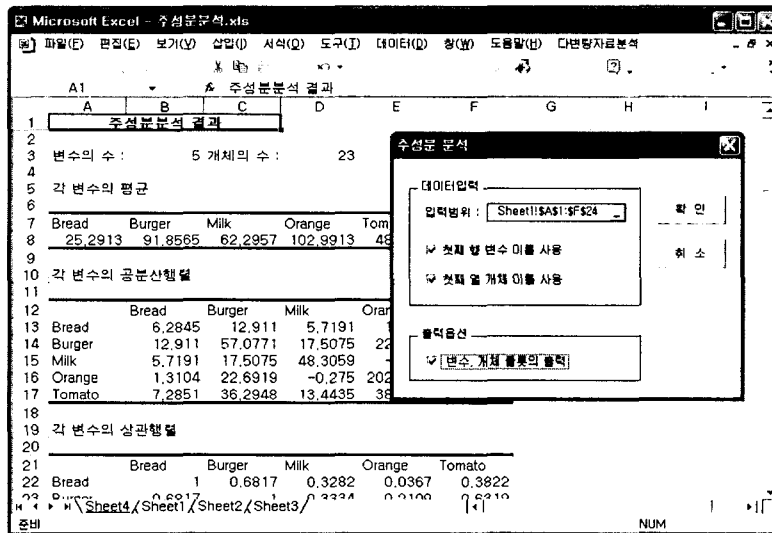


그림 3.2: 주성분분석 대화상자 및 출력결과

### 3.3. 대응분석의 수행 예

그림 3.3은 '대응분석' 대화상자 및 출력결과의 예를 보여주고 있다. 본 시스템에서 대응 분석을 수행하기 위해서는 우선 자료의 형태가 분할표 형태이어야 하는데, 만약 원자료(row data)를 가지고 있다면 먼저 Excel의 피벗테이블 작성 기능을 이용하여 데이터를 분할표 형태로 변환한 후 사용하면 된다. 대응분석 대화상자에서도 워크시트에 입력된 자료를 마우스로 드래그하거나 직접 입력범위를 지정한 후 기타 옵션들을 설정하면 원하는 결과를 얻을 수 있다.

예를 들어 '첫째 행 변수 이름 사용'과 '첫째 열 개체 이름 사용'에 체크한 다음(입력자료의 첫 번째 행과 첫 번째 열에 변수명이나 개체 이름들이 입력된 경우), 행좌표와 열좌표 옵션(디폴트는 '표준좌표')을 지정해주고 차원의 수(디폴트는 '2')를 결정한 후, 출력 옵션

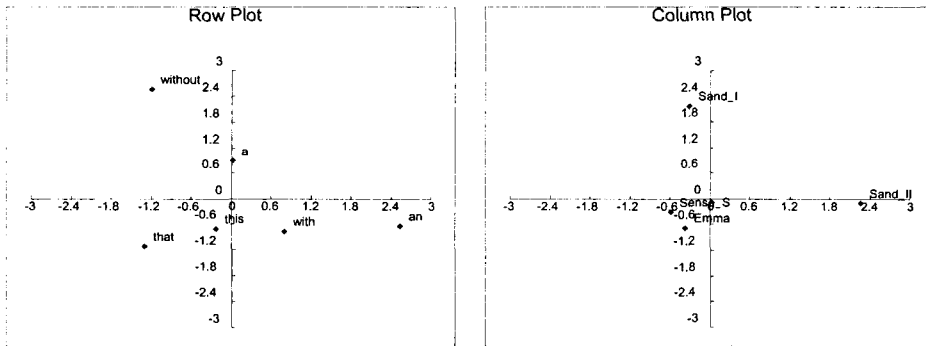
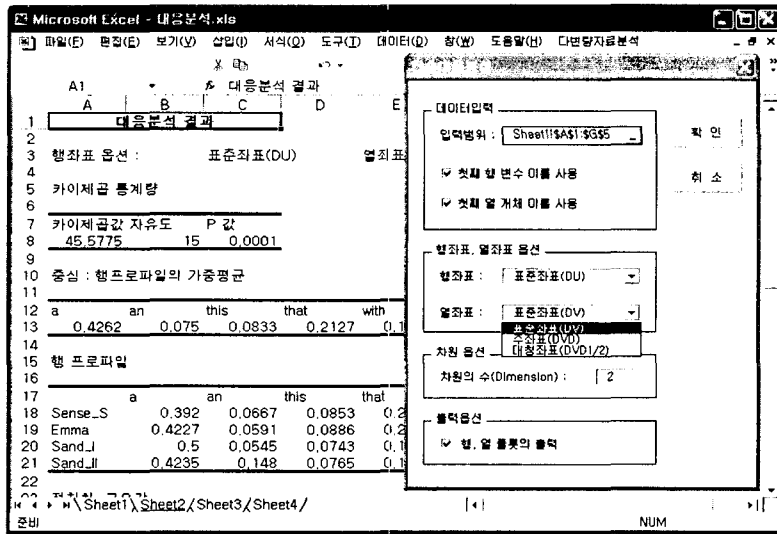


그림 3.3: 대응분석 대화상자 및 출력결과

에서 '행, 열 플롯의 출력'을 선택하고(그래프 출력을 원할 때) 확인을 클릭하면 대응분석의 결과 및 행과 열의 플롯 결과를 얻게 된다. 여기서 분석에 사용된 예제 데이터는 제인 오스틴의 세 작품과 그것의 모작인 Sandition II에서 몇 개 단어의 출현빈도에 대한 자료이다(허명희, 1999).

본 시스템에서는 그림 3.3에서와 같이 콤보박스를 이용하여 행차표 및 열차표 옵션을 선택할 수 있도록 하였으며, 디폴트는 행차표와 열차표 모두 '표준차표'이다. 또한 '차원 선택' 옵션의 디폴트는 '2'이며, 차원의 수가 2보다 작거나  $Min(\text{행의 수}-1, \text{열의 수}-1)$ 보다 크면 오류메시지 대화상자가 나타나도록 되어있다.

대응분석 대화상자를 통해 얻을 수 있는 출력결과는 카이제곱 통계량, 자유도,  $p$  값, 행프로파일 가중평균, 행프로파일, 정칙치, 고유치, 행차표, 열차표, 행렬도 등이다.

#### 4. 결론

본 연구에서는 여러 응용분야에서 가장 널리 활용되고 있는 6가지 다변량자료분석 방법을 일반 사용자가 쉽게 사용할 수 있도록 메뉴방식의 시스템을 구현하였다. 이 시스템은 Excel 매크로와 VBA를 이용하여 개발되었으며, 통합문서 소프트웨어인 Excel과 연동하여 GUI 환경에서 대화식으로 분석을 수행할 수 있다는 장점을 가지고 있다. 본 연구에서 개발된 시스템을 이용하게 되면 일반 사용자가 SAS나 SPSS 등의 고가의 소프트웨어를 구입하는데 따른 비용적 부담과 사용법을 익히는데 겪는 어려움을 크게 덜 수 있을 것이라 생각된다. 군집분석 등 기타 다변량자료분석 기법에 대한 추가, 시스템을 사용하면서 발견되는 문제점 보완 및 일반사용자들을 위한 분석기법별 대화상자에 대한 도움말 등은 차후에 지속적으로 개선해 나갈 생각이다.

#### 참고문헌

- [1] 서혜선, 김미경, 허명희 (1999). SAS AF/SCL로 구현한 다변량 수량화 시스템, <한국분류학회지>, 3, 1-11.
- [2] 조신섭, 송문섭, 이운모, 성병찬, 윤영주, 이현부 (1998). 기초통계교육을 위한 여러 통계패키지의 비교연구 및 엑셀을 이용한 한글 통계패키지의 구현, <한국통계학회 춘계 학술발표회 논문집>, 68-74.
- [3] 한상태, 강현철, 이성건, 장명석, 이덕기, 유동균. (2001). Development of Multivariate Analysis System by Using SAS/AF and SCL, *The Korean Communications in Statistics*, 8, 507-514.
- [4] 허명희 (1999). <사회과학을 위한 다변량자료분석>, 서울 : 자유아카데미.
- [5] 현기홍, 최용석 (2000). 행렬도 시스템(Biplots System)의 개발, <응용통계연구>, 13, 2, 297-306.
- [6] Codemedia 편저 (2002). *Perfect Master Excel VBA*, 서울 : 영진닷컴.

[ 2003년 7월 접수, 2003년 10월 채택 ]

## A Development of Multivariate Analysis System by Using Excel

Sang-Tae Han<sup>1)</sup> Hyuncheol Kang<sup>2)</sup> Jung-Hun Han<sup>3)</sup>

### ABSTRACT

Recently, there have been several studies to develop the multivariate data analysis system which can be readily used. The common characteristic of these studies is to develop the GUI system to which advanced statistical methods can be conveniently applied. In an extension of these studies, this study aims to supply users in various fields an interactive system with the convenience of the environment of GUI, which is constructed with the Excel macro and VBA, to apply multivariate data analysis methods easily. This system provides a graphic-oriented and menu-centered user interface in the Microsoft Excel which is widely used spreadsheet and analysis program.

*Keywords:* Multivariate Analysis System, Excel macro, VBA

- 
- 1) (336-795) Associate Professor, Dept. of Informational Statistics, Hoseo University  
E-mail: sthan@office.hoseo.ac.kr
  - 2) (336-795) Senior Lecturer, Dept. of Informational Statistics, Hoseo University  
E-mail: hychkang@office.hoseo.ac.kr
  - 3) (336-795) Graduate Student, Dept. of Informational Statistics, Hoseo University  
E-mail: cooljack94@empal.com