

Microsoft Excel의 Macro와 VBA를 이용한 다변량자료분석 시스템 개발

한상태¹⁾ · 강현철²⁾ · 이성건³⁾ · 한정훈⁴⁾

요 약

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. 이런 연구의 연장선상에서 본 연구에서는 다변량자료분석 시스템을 구현하는데 있어 사회 각 분야에서 가장 널리 활용되고 있는 사무용 프로그램인 마이크로소프트(Microsoft) Excel을 활용하여 일반 사용자들도 다변량분석을 쉽게 활용할 수 있도록 대화식 시스템을 개발하였다.

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 SAS/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, 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 '데이터분석' 기능에서 처리할 수 있는 기초통계분석이 대부분이라 할 수 있다. 즉 Excel을 기반으로 하는 다변량 자료분석 기능은 갖추고 있지 못한 문제점을 안고 있다.

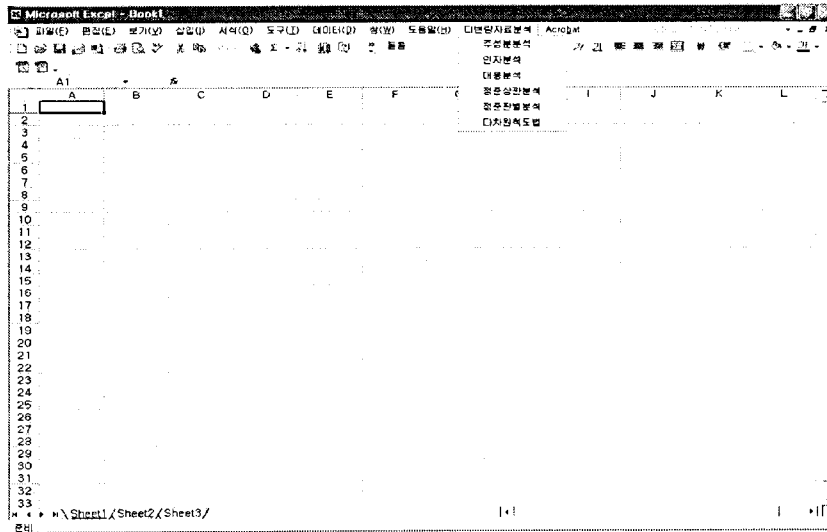
이런 제한점을 극복하기 위해 본 연구에서는 사회 각 분야에서 가장 널리 활용되고 있는 사무용 프로그램 중 하나인 마이크로소프트(Microsoft) Excel을 활용하여 다변량자료분석 시스템을 개발하고자 한다. Excel은 통합문서 프로그램으로 편리한 사용자 환경과 계산, 그래프 작성 기능 및 다양한 통계함수를 가지고 있다. 또한 기본적으로 기초통계, t-검정, 회귀분석, 분산분석 등을 '데이터분석도구'라는 추가기능에서 제공하고 있으며, '데이터분석도구'는 다른 통계패

- 1) 호서대학교 자연과학부 정보통계학전공 교수, (336-795) 충남 아산시 배방면 세출리 산 29-1
- 2) 호서대학교 자연과학부 정보통계학전공 교수, (336-795) 충남 아산시 배방면 세출리 산 29-1
- 3) 고려대학교 대학원 통계학과 박사과정, (136-701) 서울 성북구 안암동 5가 1번지
- 4) 호서대학교 대학원 수학과 통계전공 석사과정, (336-795) 충남 아산시 배방면 세출리 산 29-1

Excel을 활용한 다변량 자료분석 시스템

키지와는 달리 대화식으로 통계분석을 수행할 수 있는 장점을 가지고 있다. 그러나, Excel에서 제공하고 있는 '데이터분석도구'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 '데이터분석도구'와 같은 추가기능 파일 형태(.xla)로 다변량 자료분석 시스템을 개발하였으며, 설치방법과 사용법 또한 사용자의 편리를 위해 Excel의 '데이터분석도구' 기능과 동일하게 구성하였다. 따라서 개발된 시스템을 이용하면 일반사용자들이 굳이 통계패키지의 사용법을 익히지 않고, 쉽게 다변량자료분석을 수행 할 수 있는 장점을 제공할 수 있을 것이다.

<그림 2.1> 다변량 자료분석 시스템



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 제작 할 수는 없고, Excel, Access, Word, Powerpoint, Outlook 등의 호스트 응용프로그램을 위한 프로그래밍 언어라 할 수 있다.

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 WorkSheet와는 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 매크로구문과 Excel 기능에는 없는 고유값-고유벡터를 찾기 위한 함수 및 그래프에 레이블을 붙이기 위한 함수가 정의되어 있는 모듈을 말한다.

본 연구에서 개발한 다변량자료분석 시스템은 주성분분석(Principal Component Analysis), 인자분석(Factor Analysis), 대응분석(Correspondence Analysis), 정준상관분석(Canonical Correlation Analysis), 정준판별분석(Canonical Discriminant Analysis), 다차원척도법(Multidimensional Scaling) 등 다양한 분석을 수행할 수 있도록 되어 있다. <그림 2.1>은 다변량자료분석 시스템이 Excel에 설치되어 있는 화면이다. 각 분석 방법은 Excel의 WorkSheet에 입력된 데이터를

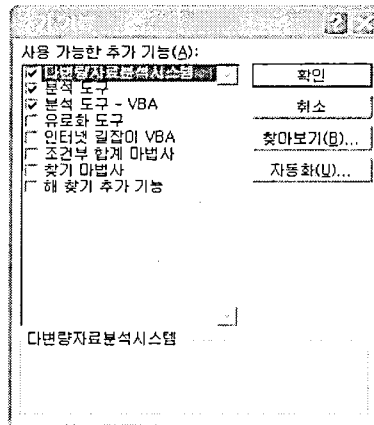
마우스로 드래그(drag)하거나 입력범위를 직접 입력함으로써 간단하게 시스템에 적용시킬 수 있고, 또한 정해진 분석단계에 의해 결과물과 그래프를 새로운 WorkSheet에 출력할 수 있도록 되어 있다.

3. 다변량자료분석 시스템 사용법

3.1 시스템 설치 및 제거방법

시스템의 설치방법은 Windows 버전에 따라 다소 차이가 있다. 설치방법에 다소 차이가 나는 이유는 Microsoft Office 프로그램의 Addin 폴더의 위치 때문이다. Windows98 버전에서는 C:\Windows\Application Data\Microsoft\AddIns 폴더에 개발된 '다변량자료분석시스템.xla'을 복사하여 붙여넣기를 한 후, Excel의 '도구→추가기능→찾아보기'에서 '다변량자료분석시스템.xla'을 선택하고 확인을 클릭하면 <그림 3.1>과 같이 추가기능 윈도우에 다변량자료분석 시스템 추가기능이 설치되는데, 여기서 확인버튼을 클릭하면 <그림 2.1>과 같이 Excel의 풀-다운 메뉴에 다변량자료분석이라는 메뉴가 생성되게 된다. 또한 Windows2000 Professional 버전과 WindowsXP 버전에서는 C:\Documents and Settings\사용자 폴더\Application Data(숨김)\Microsoft\AddIns 폴더에 '다변량자료분석시스템.xla'파일을 복사하여 붙여넣기를 한 후 Windows98 버전에서와 같은 방법으로 설치를 하면 된다. 반대로 설치된 추가기능을 해제할 때는 '도구→추가기능'에서 '다변량자료분석시스템' 체크박스의 체크(√)된 부분을 해제시키면 된다.

<그림 3.1> 추가기능 윈도우에 다변량자료분석 시스템이 추가된 모습

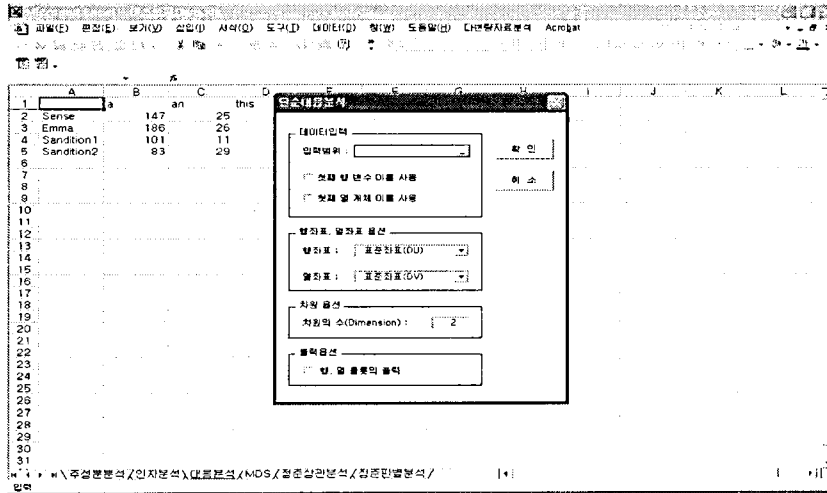


3.2 분석방법

<그림 2.1>과 같이 풀-다운메뉴가 열리고 사용자가 원하는 분석방법을 클릭하면 해당되는 다변량 분석화면이 나타나게 된다. 예를 들어 <그림 3.2>는 <그림 2.1>에서 대응분석을 클릭하여 나타난 "대응분석 대화상자"이다. 대응분석 대화상자에서 WorkSheet에 입력된 자료를 마우스로 드래그하거나 직접 입력범위를 지정하고, "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 옵션에서 "행, 열 플롯의 출력"을 선택하고(그래프 출력을 원할 때) 확인을 클릭하면 대응분석결과 및 행과 열의 플롯 결과를 얻게 된다.

Excel을 활용한 다변량 자료분석 시스템

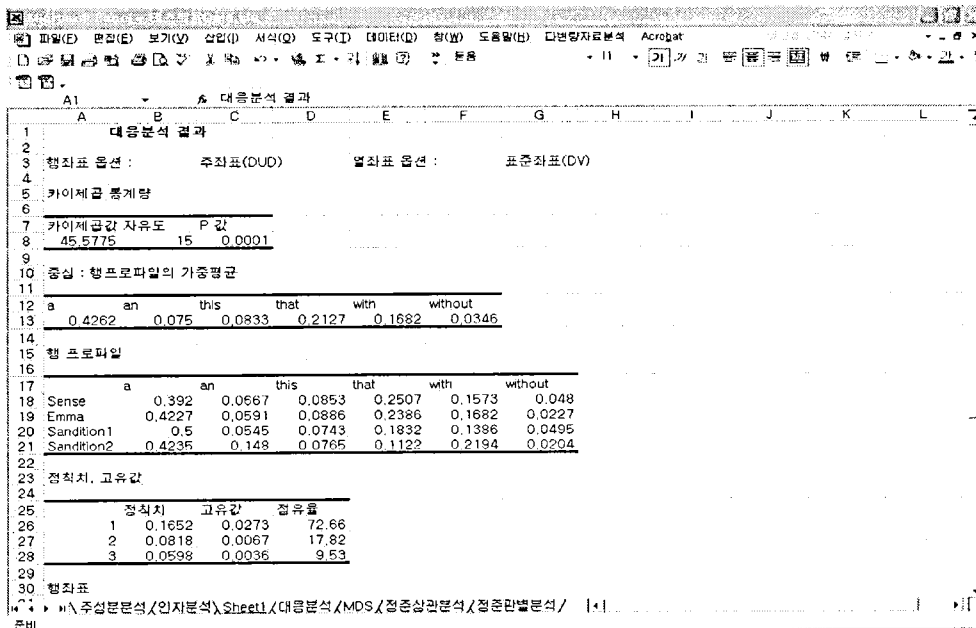
<그림 3.2> 대응분석 대화상자



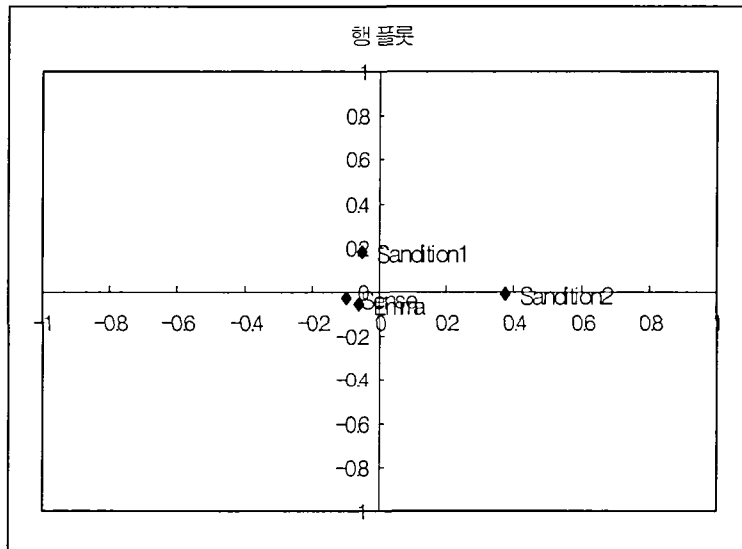
3.3 분석결과

분석결과는 <그림 3.3>과 같이 활성화되어 있던 WorkSheet 앞에 새로운 WorkSheet로 출력된다. 만약에 분석자가 “행, 열 플롯의 출력” 옵션을 선택하였다면 <그림 3.4>와 <그림 3.5>같이 Excel 차트가 결과화면 <그림 3.3>에 같이 출력된다.

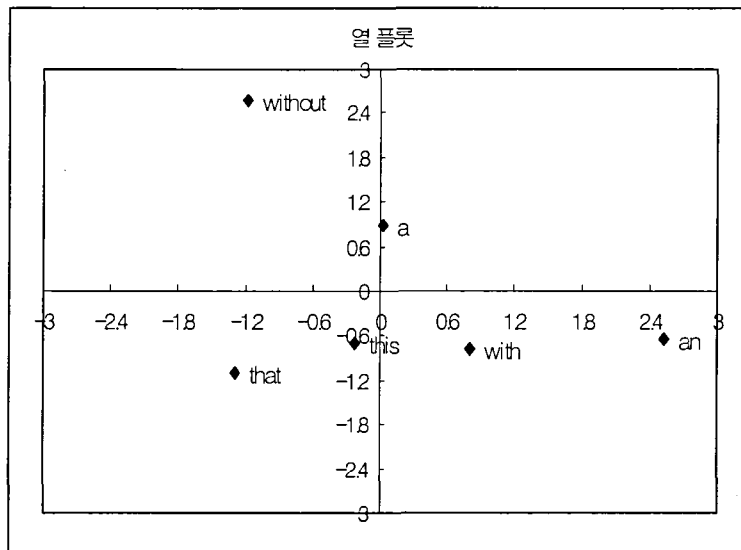
<그림 3.3> 대응분석 결과 화면



<그림 3.4> 행 플롯



<그림 3.5> 열 플롯



4. 결론

본 연구에서는 가장 널리 활용되고 있는 6가지 다변량분석방법을 일반 사용자가 쉽게 적용할 수 있도록 Excel 매크로와 VBA를 이용하여 대화식 시스템을 구현하였다. 앞 절에서는 대응분석의 예를 들어 다변량 자료분석 시스템의 기능을 설명하였는데, 다른 다변량자료분석에 대해서도 유사한 과정을 거쳐 쉽게 분석결과를 얻을 수 있다. 일반 사용자가 SAS나 SPSS 등의 고가의 소프트웨어를 구입하고, 또 사용법을 익히는데 겪는 어려움을 본 연구에서 개발된 시스템을 통해 조금이나마 덜 수 있을 것이라 생각된다. 본 시스템에서 아직 개발하지 못한 군집분석

Excel을 활용한 다변량 자료분석 시스템

및 각 분석기법 대화상자에 대한 도움말 등은 차후에 지속적으로 개발하여 시스템에 추가 시켜 나갈 생각이다.

참고문헌

- [1] 서혜선·김미경·허명희 (1999), SAS AF/SCL로 구현한 다변량 수량화 시스템, 「한국분류학회지」, 제3권, 1-11.
- [2] 조신섭·송문섭·이윤모·성병찬·윤영주·이현부 (1998), 기초통계교육을 위한 여러 통계패키지의 비교연구 및 엑셀을 이용한 한글통계패키지의 구현, 「한국통계학회 춘계 학술발표회 논문집」, 68-74.
- [3] 한상태·강현철·이성진·장명석·이덕기·유동균 (2001), Development of Multivariate Analysis System by Using SAS/AF and SCL, *The Korean Communications in Statistics, Vol. 8*, 507-514.
- [4] 허명희 (1999), 『다변량 수량화』, 서울 : 자유아카데미.
- [5] 현기홍·최용석 (2000), 행렬도 시스템(Biplots System)의 개발, 「응용통계연구」, 제13권, 297-306.
- [6] Codemedia 편저 (2002), Perfect Master Excel VBA, 서울 : 영진닷컴.