

Excel VBA를 이용한 행렬도 시스템의 구현

유성모¹⁾, 서용환²⁾

요 약

자료행렬에서 개체와 변수간의 관계성을 시각적으로 표현하기 위한 방법 중의 하나가 행렬도이다. 본 논문에서는 전문적인 통계 패키지를 이용한 행렬도 구현이 아니라, 가장 널리 사용되는 응용프로그램 중의 하나인 Excel 에서 VBA를 이용하여 행렬도 시스템을 구현하였다.

주요용어 : 밀집도, 행렬도, Excel, VBA

1. 서론

행렬도(Biplot)는 자료행렬에 대하여 행그림과 열그림을 하나의 그림에 동시에 나타내어 분석하는 방법이다. 삼차원 이상의 자료행렬에 대해서는 차원축소가 필요하며, 차원축소에 대한 이론적인 근거는 자료행렬에 대한 비정칙치분해의 이해로부터 시작된다.

본 논문은 일반적으로 가장 많이 사용되는 응용프로그램인 Excel 상에서 VBA(Visual Basic for Application)를 이용하여 행렬도 시스템을 구현하는 것에 목적을 두고 있다. 별도의 설치과정 없이, Excel의 추가기능을 이용하여 간단히 행렬도를 작성할 수 있다. 이 행렬도 시스템은 주성분 행렬도, 요인 행렬도를 작성하며, 밀집도, 근사적합도, 비정칙치 분해에 대한 결과 값들을 Excel 상의 워크시트에 출력한다.

2. Excel 상에서의 행렬도 시스템 구현.

2.1. 행렬도(Biplot)의 이해

행렬도에서는 자료행렬의 개체(행)와 변수(열)의 관계를 저차원상의 그림으로 나타낸다. 자료행렬 $X_{n \times p}$ 는 두 행렬($G_{n \times 2}, H_{2 \times p}$)로 분해 가능하며, 이는 $G_{n \times 2}$ 와 $H_{2 \times p}$ 의 곱으로 근사된다. 또한, 자료행렬에 대한 비정칙치분해로부터 대각행렬을 어느 행렬에 곱하는가에 따라 G 와 H 의 행렬형태는 달라지게 된다.

$$X_{n \times p} = U_{n \times p} D_{\sqrt{\lambda_p}}^\sigma D_{\sqrt{\lambda_p}}^1 V_{p \times p} = G_{n \times p} H_{p \times p} \simeq G_{n \times 2} H_{2 \times p}$$

여기서, U 와 V 는 각각 $UU^T = I_p$ 와 $VV^T = I_p$ 를 만족하는 직교행렬이고, D 는 자료행렬의 공분산행렬 $S = \frac{1}{n-1}(X-\mu)'(X-\mu)$ 에 대한 고유치들의 제곱근들을 원소로 하는 대각

1) (339-700) 충남 연기군 조치원읍, 고려대학교 정보통계학과 부교수.
E-mail : syoo@korea.ac.kr

2) (339-700) 충남 연기군 조치원읍, 고려대학교 정보통계학과 석사과정.
E-mail : windwal@naver.com

Excel VBA를 이용한 행렬도 시스템의 구현

행렬이다. $\alpha=1$ 인 경우를 주성분 행렬도라 하며, G행렬을 통해서 개체간의 유클리드(Euclid) 거리를 파악할 수 있다. $\alpha=0$ 인 경우를 요인행렬도라 하며, 중심화된 자료행렬을 이용할 경우, G행렬을 통해서 개체간의 마할라노비스(Mahalanobis) 거리를 구할 수 있으며, H 행렬을 통해서 변수간의 상관관계를 파악할 수 있다. (허명희 1993, v-8 참조)

2.2. 개체와 변수간의 밀접도

표준화된 자료행렬에 대한 요인행렬도에서 i 번째 개체(행)와 j 번째 변수(열)에 해당하는 표준화된 자료 x_{ij} 를 구성하는 두 벡터의 사이각의 코사인을 i 번째 개체(행)와 j 번째 변수(열)의 밀접도 r_{ij} 로 정의할 수 있다. 즉,

$$r_{ij} = \cos \theta_{ij} = \frac{\vec{g}_i \cdot \vec{h}_j}{|\vec{g}_i| |\vec{h}_j|} \quad (\text{단, } i=1, \dots, n; j=1, \dots, p; 0^\circ \leq \theta_{ij} \leq 180^\circ)$$

이다. (유성모, 김상우 2001)

2.3. 행렬도 시스템의 구현

다음은 행렬도 시스템을 이용하여, 각각 15대/16대 국회의원 선거결과에 대한 요인행렬도이다. 지역성을 나타내는 G행렬 벡터는 행렬도의 좌표 위에 직역을 나타내는 레이블로 표시하였고, 정당을 나타내는 H행렬 벡터는 원점과 좌표를 연결하는 직선으로 표시하였다.

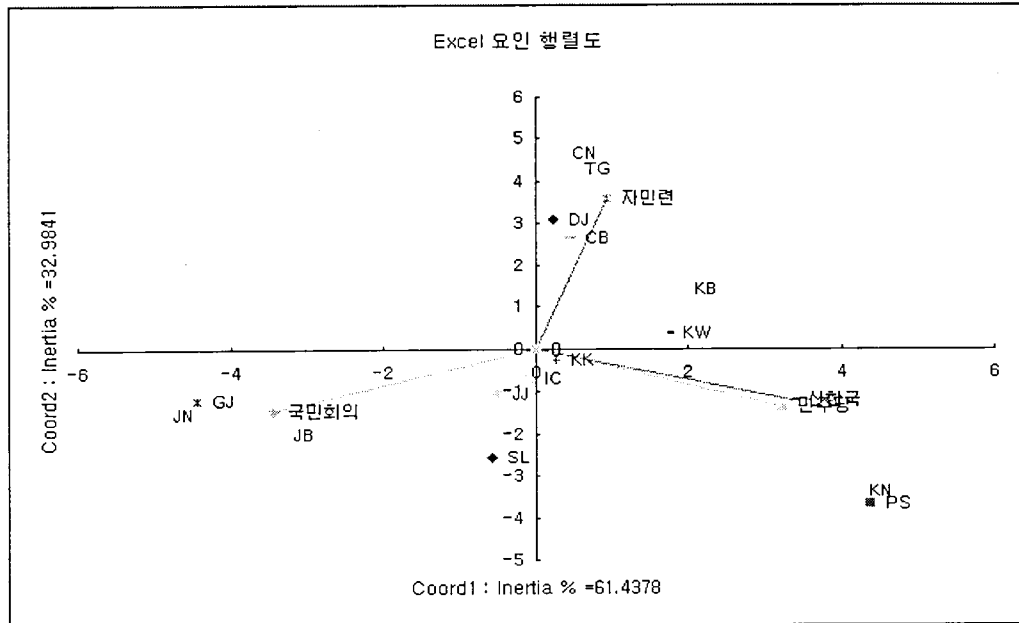


그림 2.1 : 15대 국회의원 선거결과에 대한 요인행렬도

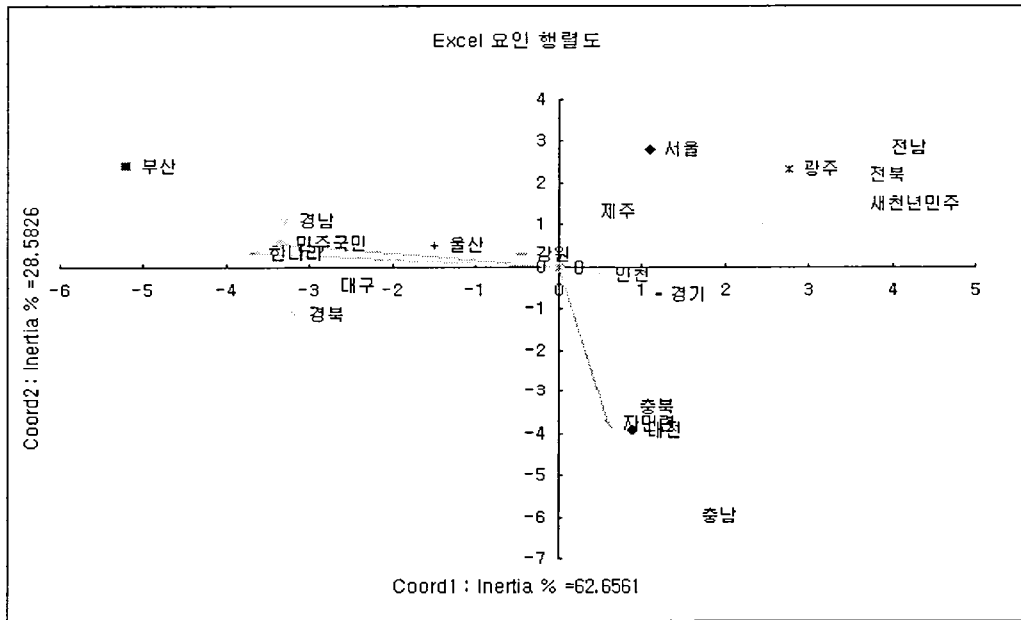


그림 2.2 : 16대 국회의원 선거결과에 대한 요인행렬도

그림 2.1과 그림 2.2에서의 지역간 거리는 마할라노비스 거리를 나타낸다. 거리가 가까울수록 지역간의 유사성이 높음을 나타내고, 멀수록 지역간의 유사성은 낮음을 의미한다.

2.4. 프로그램의 배포 및 설치

프로그램은 <http://www.kuinfo.net/bbs.asp?index=ysm> 에서 다운로드할 수 있다. Excel 추가기능 파일(XLA)의 형식으로 되어 있으며, Excel 의 "추가기능 관리자"를 이용하여 설치한다.

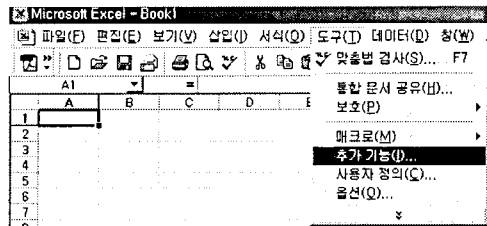


그림 2.3 추가기능 위치

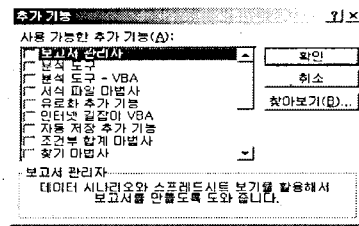


그림 2.4 XLA 설치

상단의 메뉴에 "행렬도 시스템"이라는 메뉴에서, "행렬도 시작하기"를 클릭하면, 다음과 같은 행렬도 시스템 폼이 화면에 표시된다.

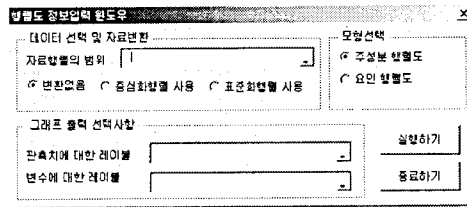


그림 2.5 행렬도 시스템 폼

폼상에서 자료행렬에 대한 범위, 변환형태, 모형선택 등을 한 다음, 실행을 하면 "Output" 워크시트가 추가되면서, 결과가 표시된다.

실행에 대한 자세한 설명은, 도움말을 참조하도록 한다.

3. 결 론

본 논문에서는 Excel VBA를 이용하여 행렬도 시스템을 구현하였다. 이 행렬도 시스템은 주 성분 행렬도, 요인 행렬도, 밀접도 등에 대한 결과를 출력한다. 그리고, 전문적인 통계 패키지를 사용하지 않고, 행렬도에 대한 쉬운 접근 방안을 제시하였다.

참고문헌

- [1] 유성모, 김상우 (2001). 행렬도를 이용한 개체와 변수간의 밀접도에 대한 연구, <응용통계연구>, 제 14권 2호, pp.393-399.
- [2] 허명희 (1993). *통계상담의 이해*, 자유아카데미, 서울.
- [3] Roman, Steven. (1999). *Writing Excel Macros*, O'Reilly.
- [4] Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (1992). *Numerical Recipes in C*, CAMBRIDGE University Press.