

## 행렬도 시스템(BIPLOTS SYSTEM)의 개발 \*

최용석<sup>1)</sup> 현기홍<sup>2)</sup>

### 요약

행렬도(biplots)란 행과 열을 그래프에 동시에 나타내어 이들의 관계를 살피려는 다변량 그래픽분석 기법이다. 이를 활용하기 위해서 사용자들이 수작업으로 일일이 프로그래밍해야 했던 불편함에서 벗어나 GUI(Graphic User Interface) 환경으로 메뉴에 의해서 사용하기에 편한 행렬도 시스템의 개발을 소개한다. 구체적으로 본 시스템의 개발 방법, 과정, 구성 그리고 예제를 통해 실행결과를 확인하고 결론을 통해 추후 연구과제와 발전방향을 제시한다.

주요용어: 행렬도, GUI, SAS AF/SCL.

### 1. 서론

통계적 자료처리를 위하여 가장 널리 사용되고 있는 SAS는 많은 모듈로 구성되어 있고 수많은 처리기능을 지니고 있으나, 사용법의 숙지와 결과의 올바른 해석을 위하여 적지 않은 통계적 이론과 노력이 필요하다. 더군다나 행렬도와 같은 다변량 그래픽분석 기법의 활용에 대한 중요성은 날로 커져가고 있으며, 이를 위한 이론적 연구도 활발히 수행되고 있다. 또한 이러한 기능들을 기존의 SAS 모듈에 추가하여 기존 기능과 연계하여 처리할 수 있는 시스템의 개발은 사용자들의 편의는 물론 비용, 효과 면에서 매우 효율적이라 판단된다.

본 연구의 목적은 크게 두 가지이다. 첫째, 기존의 SAS/IML을 이용하여 행렬도의 알고리즘을 일일이 프로그래밍해서 처리해야 하는 불편함에서 벗어나 사용자 중심의 GUI 환경으로 구축하여 편의성을 제공하자는 것이다(전홍석, 김승환, 1996).

둘째, 하드웨어와 소프트웨어의 발전에 힘입어 데이터로부터 유용한 정보를 얻기 위한 여러 시스템의 구축이 확산되고 있는 정보사회의 시대적 흐름을 반영하자는 것이다. 근래에 들어 컴퓨터의 보급이 점점 확산되면서 컴퓨터를 활용한 시스템 개발 및 자료분석에서 활발한 연구가 이루어지고 있고 정보화 사회가 그러한 연구를 더욱 부추기고 있는 듯하다. 또한 정보사회에서 수없이 쏟아지는 대규모 데이터를 데이터 웨어하우징으로 불리는 대규모 데이터베이스를 구축하고 있고, 이로부터 유용한 정보를 추출하기 위하여 여러 시스템이 구축되고 있는 실정이다.

특히, 통계학 이론을 바탕으로 한 시스템 개발의 국내 사례로 허문열(1995), 유종영 외 2인(1997), 서혜선 외 2인(1999) 등이 있다.

\* 행렬도 시스템의 사용을 원하는 경우 <http://home.pusan.ac.kr/~yschoi>를 접속하면 된다.

1) (609-735) 부산시 금정구 장전동, 부산대학교 자연과학대학 통계학과, 부교수

E-mail: yschoi@hyowon.pusan.ac.kr

2) (609-735) 부산시 금정구 장전동, 부산대학교 자연과학대학 통계학과, 식사과장

E-mail: shoma@lge.com

## 2. 시스템 개발 방법 및 과정

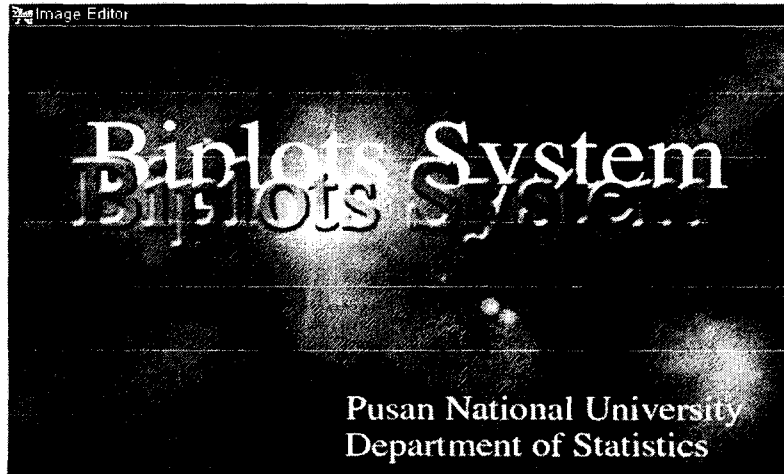


그림 2.1: 로고 화면

본 시스템은 사용자 중심적이고 그래픽 위주인 GUI 환경으로 구현하기 위해 객체지향적 비주얼도구인 SAS/AF와 화면제어를 위해 사용되는 언어인 SAS/SCL로 개발하였다. 시스템에서 나타난 각 프레임들은 SAS/PROC BUILD 절차에서 디자인되었다. 그리고 세부적으로 SAS의 여러 다른 제품들(IML, MACRO, GRAPH, INSIGHT, ASSIST)을 이용하여 통계계산 및 시스템환경을 구성하였다(SAS Institute Inc., 1993a, 1993b, 1994). 시스템은 다음의 5단계 과정을 통하여 개발되었다.

### 단계 2.1 시스템 환경설정 파일 제작

행렬도 시스템이 구동하기 위한 환경설정 파일 작성과 로고 화면인 그림 2.1 및 주 메뉴 프레임인 그림 2.2를 디자인한다.

- (1) 행렬도 시스템이 실행되기 위한 환경설정 파일인 config.sas를 SAS 패키지의 환경설정파일을 참고하여 작성한다.
- (2) 그림 2.1, 그림 2.2는 이미지 디자인 전문 프로그램을 이용하여 제작되었고, 행렬도 시스템이 구동될 때 자동으로 화면에 나타나게 된다.
- (3) 그림 2.2의 각 객체가 선택되었을 때 실행될 수 있게 SAS/SCL로 프로그래밍 한다.

### 단계 2.2 자료 처리를 위한 프레임 디자인

본 시스템은 SAS 패키지 내에서 구동되므로 외부파일을 SAS Data Set으로 변환시키는 작업과 수정 및 편집할 수 있는 프레임이 필요하다.

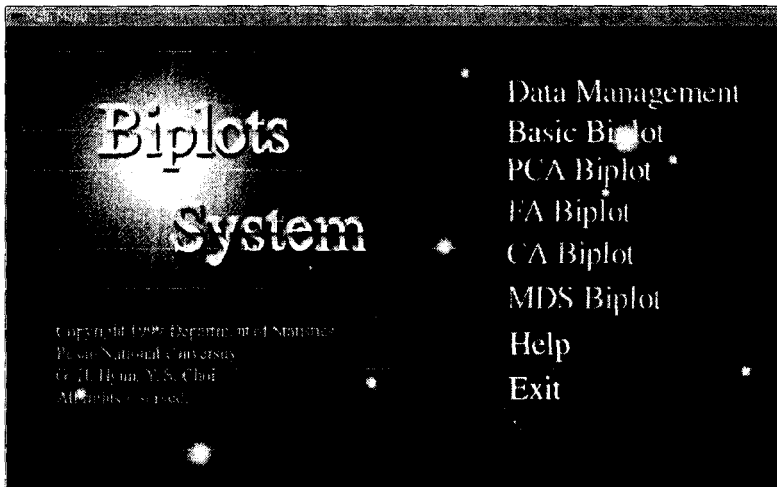


그림 2.2: 주 메뉴 프레임

- (1) 자료 처리 프레임은 외부파일을 SAS Data Set으로 변환하는 부분, SAS/ASSIST 제품을 활용하여 기존의 SAS Data Set을 수정 및 편집하는 부분, 다른 프레임들과 마찬가지로 도움말 기능과 이전 프레임으로의 복귀를 위한 객체들로 디자인되어 있다.
- (2) 각 객체가 선택되었을 때 실행될 수 있게 SAS/SCL로 프로그래밍 한다.

### 단계 2.3 각 행렬도의 옵션 기능을 가진 프레임 디자인

본 시스템은 기본적으로 GH, SYM, JK 세 가지 행렬도를 제공한다. 또한 PCA(주성분 분석), FA(인자분석), CA(대응분석), MDS(다차원척도)도 넓게는 행렬도의 범주에 포함시킬 수 있으므로 위의 각 행렬도가 필요로 하는 옵션 기능들을 디자인한다(최용석, 1999, p. 3).

- (1) 세 가지 기본 행렬도 옵션 프레임은 그림 2.4와 같이 SAS Data Set을 불러오는 객체, 행과 열 변수를 각각 선택하는 객체, Scaling Type, Biplot Type, Dimension을 결정할 수 있는 객체들로 디자인되어 있다.
- (2) 각 객체가 선택되었을 때 실행될 수 있게 SAS/SCL로 프로그래밍 한다.

### 단계 2.4 각 행렬도의 출력결과 프레임 디자인

각 행렬도의 출력결과는 통계계산 결과와 플롯인데 MDS 행렬도를 제외하고 모두 같은 프레임으로 디자인되어 있다.

- (1) 출력결과 프레임에서는 통계계산 결과와 플롯을 보여주는 객체와 저장 및 인쇄 등의 툴바 객체를 디자인한다.

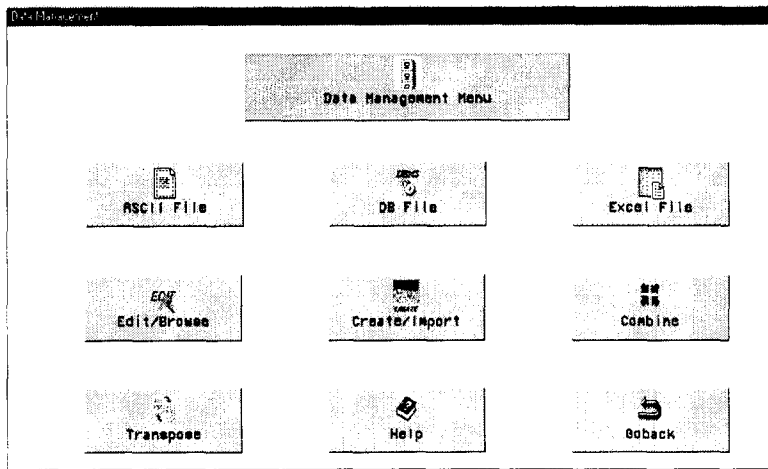


그림 2.3: 자료 처리 프레임

(2) 각 객체가 선택되었을 때 실행될 수 있게 SAS/SCL로 프로그래밍 한다.

### 단계 2.5 도움말 프레임 디자인 및 컴파일

이 단계에서는 지금까지 단계 2.1부터 단계 2.4까지 프레임을 디자인하면서 도움말 기능을 제공하는 Help 객체에 대한 프레임을 각각 디자인한다. 최종적으로 모든 프레임 파일과 SCL 파일을 컴파일해서 연결시키고, 필요에 따라 수정한다.

## 3. 행렬도 시스템의 구성

이 장에서는 행렬도 시스템의 구성에 대해 알아본다. 주된 구성은 자료 처리를 할 수 있는 Data Management, 세 가지 기본 행렬도(GH, SYM, JK)를 제공하는 Basic Biplot, 그 밖의 행렬도인 PCA Biplot, FA Biplot, CA Biplot, MDS Biplot, 마지막으로 도움말 기능을 가진 Help와 시스템 종료를 위한 Exit로 구성되어 있다.

### 3.1. 자료 처리(DATA MANAGEMENT)

자료처리 프레임은 행렬도 시스템에서 필요로 하는 SAS Data Set의 생성과 관리를 위한 여러 객체들로 구성되어 있다.

#### (1) ASCII File, DB File, Excel File

- 외부파일을 SAS Data Set으로 변환시키는 기능을 가진 객체들로 예를들어 그림 3.1은 아스키 파일을 SAS Data Set으로 변환시키는 기능을 가진 프레임이다.

#### (2) Edit/Browse, Create/Import, Combine

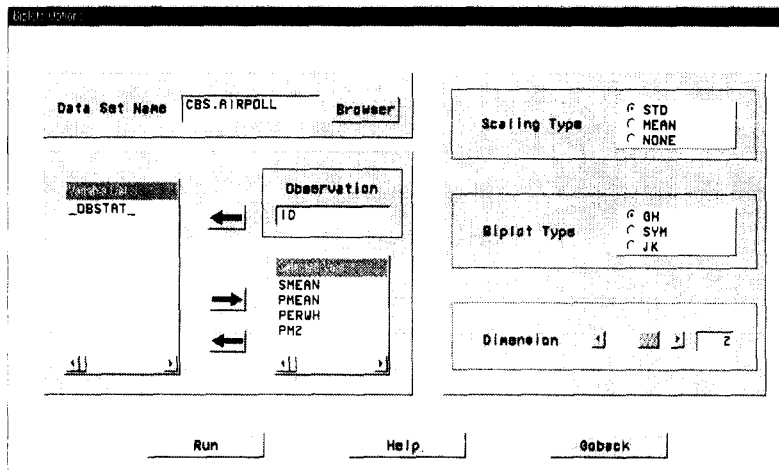


그림 2.4: 기본 행렬도 옵션 프레임

- SAS Data Set을 수정 및 편집할 수 있는 기능을 가진 객체로 SAS/ASSIST 제품을 활용하고 있다.

### 3.2. 세 가지 기본 행렬도

행렬도는 기본적으로 세 가지가 있는데 그것은 GH, SYM, JK 행렬도이며, 본 시스템에서 Basic Biplot으로 구성되어 있다.

- (1) 그림 2.4 - 기본 행렬도 옵션 프레임으로 자료를 불러들이기 위한 Browser 객체, 자료의 변수들을 나열하는 List Box Attributes 객체, 행렬도의 Scaling Type, Biplot Type, Dimension을 결정할 수 있는 객체들로 구성되어 있다.
- (2) 그림 4.1 - 행렬도 통계계산 결과 프레임으로 실행결과를 보여주는 Catalog Entry Viewer Attributes 객체와 저장 및 인쇄기능을 가지는 Toolbar Attributes 객체로 구성되어 있다.
- (3) 그림 4.2 - 행과 열을 동시에 플롯하는 프레임으로 그래프를 보여 주는 SAS/GRAPH Output Attributes 객체와 인쇄기능, 이전 화면으로 돌아가기 등을 가지는 Toolbar Attributes 객체로 구성되어 있다.

### 3.3. 그 밖의 행렬도

행렬도는 넓은 의미에서 PCA, FA, CA, MDS를 행렬도에 포함시킬 수 있다.

#### (1) PCA, FA, CA, MDS 행렬도

- 이들 행렬도에 대한 프레임에서 볼 수 있는 각 옵션들은 SAS/STAT 제품의 각 절차에서 사용하는 옵션들을 그대로 사용하고 있다.

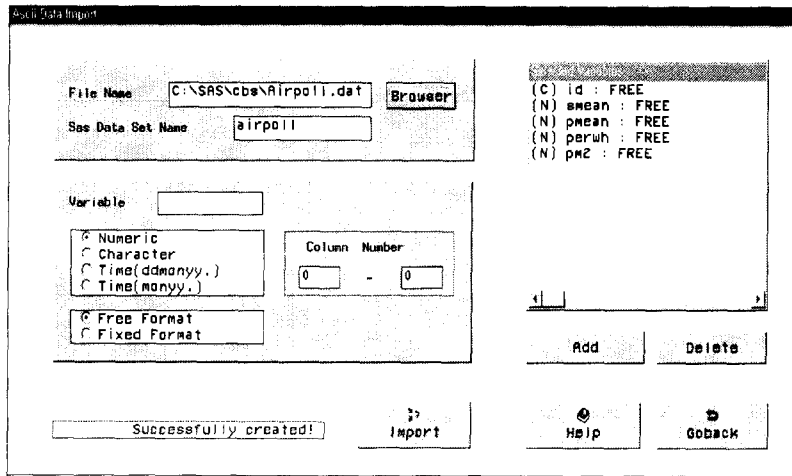


그림 3.1: 아스키 자료 불러오기 프레임

## (2) 통계계산 결과 및 플롯 프레임

- 기본 행렬도에서 사용하는 프레임과 동일한 프레임을 사용하고 있다.

## 3.4. 도움말(HELP) 및 시스템 종료(EXIT) 기능

도움말 기능은 필요에 따라 각 프레임에 선택적으로 포함되어 있다. 그리고 Exit는 행렬도 시스템을 종료하기 위한 객체로 확인 프레임이 나오면서 사용자에게 종료여부를 묻게 된다.

## 4. 자료를 이용한 시스템의 실행 예

여기서는 표 4.1의 실제 자료를 사용하여 본 시스템의 실행 과정을 살펴본다. 이 자료는 10개 지역의 공기오염(air pollution)에 관련되어 있는 네 가지 변수에 의해서 측정된 자료로써 측정변수 네 가지는 SMEAN(평균 황화물), PMEAN(평균 미립자), PERWH(백인 비율), PM2(평방마일내 인구밀도)이다(Jobson, 1992, p. 348).

우선 SAS Data Set을 영구히 저장할 수 있는 저장소인 라이브러리를 등록하는 일이 필요한데, 본 시스템이 cbs라는 라이브러리를 자동으로 생성해 주므로 cbs에서 SAS Data Set을 관리할 수 있다. 다음으로 공기오염자료가 아스키형태로 저장되어 있으므로 외부파일을 SAS Data Set으로 변환시키기 위해 주 메뉴 프레임에서 Data Management를 클릭한다. 그러면 그림 2.3의 여러 객체들 중에서 아스키 파일을 SAS Data Set으로 변환시킬 수 있는 ASCII File 객체를 클릭한다. 그림 3.1이 나타나면 브라우저를 통해 외부파일을 불러들여서 변수명, 변수형태, 열 번호 등을 기입한 다음 Import객체를 클릭한다. 만일 SAS Data Set이 성공적으로 생성되었다면 성공을 알리는 메시지가 나타나면서 cbs 라이브러리에 자료파일이 생성된다.

표 4.1: 공기오염자료

지역	SMEAN	PMEAN	PERWH	PM2
a	37	108	96.8	49.3
b	80	112	87.3	52.4
c	50	244	96.7	22.5
d	56	78	63.5	22.9
e	119	92	93.3	32.1
f	69	125	82.4	26.6
g	128	114	95.3	93.2
h	60	99	70.7	19.6
i	47	76	83.4	25.7
j	41	81	92.5	51.1

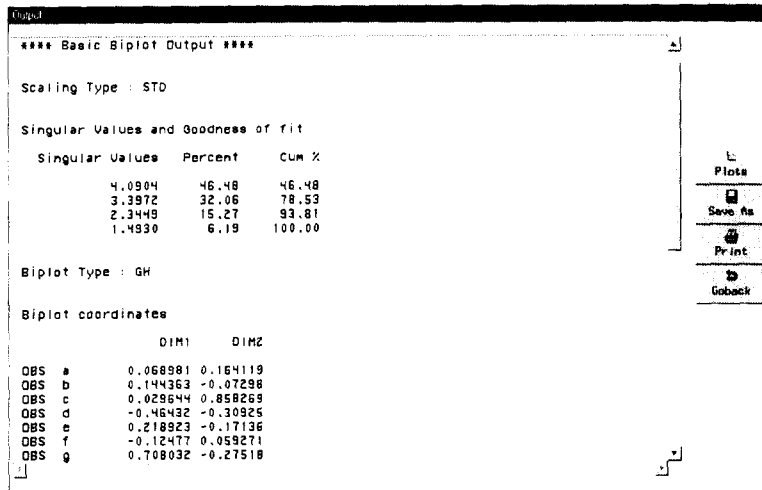


그림 4.1: 통계계산 결과 프레임

다음으로 여러 행렬도 중에서 기본 행렬도를 실행하여 본다. 주 메뉴프레임에서 Basic Biplot을 클릭하면 기본 행렬도 옵션 프레임이 나타난다. 여기서 브라우저를 통해 그림 2.4와 같이 필요한 SAS Data Set을 불러들이고, 행렬도에서 행과 열에 해당하는 변수이름과 Scaling type, Biplot Type, Dimension 등을 각각 선택하여 Run을 클릭해서 실행시킨다. 그러면 그림 4.1과 같은 출력결과를 볼 수 있고, 필요에 따라 Save As와 Print 객체를 클릭하여 파일로 저장하거나 및 프린터로 출력할 수 있다. 또한 플롯을 보기 원한다면 Plots 객체를 클릭하여 그림 4.2에 나타난 것처럼 행과 열을 동시에 볼 수 있다.

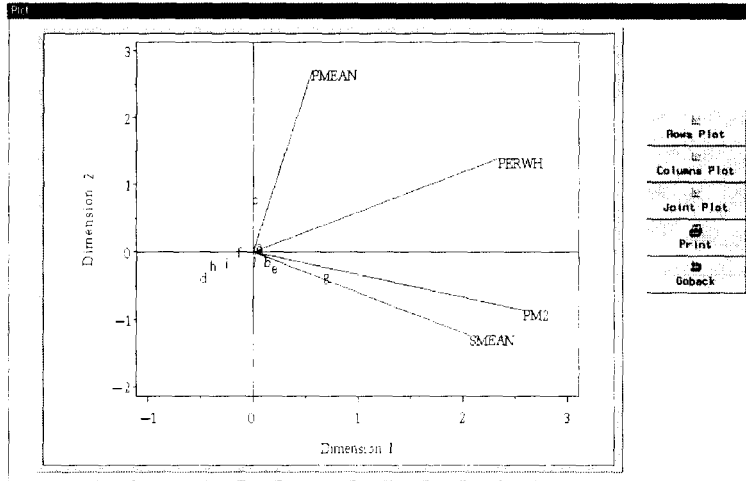


그림 4.2: 행과 열 플롯 프레임

## 5. 결론

대부분의 사용자는 SAS의 방대한 절차 및 언어를 어려워하는 것이 현실이다. 그래서인지 SAS는 Version 6.08 이후 윈도우 환경으로 바뀌면서 SAS/ASSIST, SAS/INSIGHT 등 그래픽 위주의 사용자 환경과 메뉴방식의 SAS 프로그램 개발로 발전되어 왔다. 본 논문도 다변량 그래픽 분석기법의 하나인 행렬도를 사용자들이 기존에 느꼈던 수작업의 불편함에서 벗어나 GUI 환경에서 통계계산 결과 및 그래프를 편리하고 쉽게 확인할 수 있는 시스템이다. 즉 세 가지 기본 행렬도와 PCA, FA, CA, MDS 행렬도의 출력결과와 그래프를 확인할 수 있고 출력 결과를 파일로 저장할 수 있어 추후 다른 분석에도 이용할 수 있다.

본 시스템은 SAS Version 6.12에 초점을 맞추었고 운영체제는 WINDOWS 95, 98에서만 수행되게 설계되었으며, 해상도는 800 × 600을 갖는 화면을 기준으로 제한되어 구축된 것이 기술적인 문제이다. 추후 연구과제로는 최근 인터넷 환경과 연계해서 웹(web) 기반에서 구동되는 행렬도 시스템을 고려할 수 있고, 출력결과에서 자료의 일부를 변화시키거나 플롯의 형태를 변화시켰을 때 순간적으로 플롯이나 자료에 어떠한 영향을 주는지를 즉시 알 수 있는 동적 그래픽스(dynamic graphics) 기법을 이용한다든지, 모형의 진단을 평가할 수 있는 그래픽 기법 등을 제시 해 준다면 자료를 탐색적으로 분석하는데 보다 뛰어나고 편리한 시스템이 되리라 생각된다.



## 참고문헌

- [1] 서혜선, 김미경, 허명희 (1999). SAS AF/SCL로 구현한 다변량 수량화 시스템, < 한국 분류학회>, 제 3권, 1-11.
- [2] 유종영, 안기수, 허문열 (1997). 동적 그래픽스에 의한 회귀진단 시스템(REDS)의 구현, <응용통계연구>, 제 10권, 2호, 241-251.
- [3] 전홍석, 김승환 (1996). <SAS를 활용한 윈도우 프로그래밍 -SAS/AF와 SCL->, 자유아카데미, 서울.
- [4] 최용석 (1999). <행렬도의 이해와 응용>, 부산대학교 출판부, 부산.
- [5] 허문열 (1995). 컴퓨터 그래픽스에 의한 이원 분산분석. <응용통계연구>, 제 8권, 1호, 75-87.
- [6] Jobson, J.D. (1992). *Applied Multivariate Data Analysis*, Springer-Verlag, New York.
- [7] SAS Institute Inc. (1993). *SAS/AF Software: Frame Entry Usage and Reference*, Version 6, First Edition, SAS Institute Inc., NC: Cary.
- [8] SAS Institute Inc. (1993). *SAS Companion for the Microsoft Windows Environment*, Version 6, First Edition, SAS Institute Inc., NC: Cary.
- [9] SAS Institute Inc. (1994). *SAS Screen Control Language: Reference*, Version 6, Second Edition, SAS Institute Inc., NC: Cary.

[ 2000년 4월 접수, 2000년 6월 채택 ]

## The Development of Biplots System \*

Yong-Seok Choi <sup>1)</sup> Gee-Hong Hyun <sup>2)</sup>

### ABSTRACT

Many users have made the most often use of the SAS system in statistical data analysis all over the world. But it is difficult to use the grand procedures and language of the SAS system. Therefore the side of program development has changed in the graphic-oriented and menu-centered way like SAS/ASSIST, SAS/INSIGHT after a version 6.08 turned into the Window environment. A biplots is a multivariate data analysis technique that graphically describes both relationships among the multidimensional observations and relationships among the variables. But there were not the procedure and graphic interface for a biplots algorithm in the SAS system. In this paper, there are two objects. First, we supply users with the convenience of the environment of GUI, which is constructed with SAS/AF and SCL, to solve the problem that we have programed a biplots algorithm with the SAS/IML one by one. Second, we reflect the current of the Information Age which means the spread of various kinds of system construction to extract useful information from data with the help of the development of hardware and software.

*Keywords:* Biplots; GUI; SAS AF/SCL.

---

\* If you want to use this system, contact <http://home.pusan.ac.kr/~yschoi>.

1) Associate Professor, Department of Statistics, Pusan University. E-mail: yschoi@hyowon.pusan.ac.kr

2) Graduate, Department of Statistics, Pusan University. E-mail: shoma@lge.com