

행렬도를 이용한 대학 신입생의 진로의식 분석

최혜미¹ · 박찬용² · 이상협³ · 정성석⁴

¹전북대학교 통계학과, ²전북대학교 통계학과, ³전북대학교 통계학과, ⁴전북대학교 통계학과

(2010년 3월 접수, 2010년 8월 채택)

요약

행렬도는 고차원의 자료를 저차원 공간에 투영하여 자료를 시각화하는 비교적 현대적인 방법으로써, 자료의 산포도, 집단 구분, 변수사이의 상관관계 등 유용한 정보들을 제공한다. 본 연구에서는 이러한 행렬도를 간략하게 소개하고, 행렬도의 구현을 위해 대중성이 높아지고 있는 무료 소프트웨어인 R의 BiplotGUI 패키지를 사용하였다. 그리고 전북대학교에서 2009년도에 실시된 신입생의 진로의식 조사 자료를 이용하여, 신입생의 선호직업과 진로성숙도의 관계를 행렬도 분석방법으로 살펴보았다.

주요용어: 행렬도, R패키지, 주성분분석, 진로성숙도.

1. 서론

행렬도는 다변량 자료를 시각적으로 알아보기 쉽게 이차원 혹은 삼차원 그래프로 표현하는 분석기법으로 Gabriel (1971)에 의해서 제안되었다. 이후 Gower와 Hand (1996)는 Gabriel (1971)의 전통적인 행렬도를 산점도의 확장된 개념으로 발전시켰다. 산점도가 해석하기 위해 많은 노력이 필요하지 않다는 장점을 가지고 있듯이, 행렬도는 통계를 모르는 사람들도 쉽게 접근할 수가 있다. Minitab, SPSS, SAS 등 다양한 통계패키지들이 행렬도를 제공하고 있으나, 기능이 제한적이고 원하는 결과를 얻는데 어려움이 따른다. 좀 더 기능적으로 뛰어난 패키지로는 행렬도를 목적으로 만들어진 XLS-Biplot, GGE-Biplot, BiPlot 등이 사용되고 있으나, 본 논문에서는 무료로 사용할 수 있으며 대중성이 높아지고 있는 R을 이용하여 행렬도 구현 방법을 소개하고 있다. 또한 사례 연구로 신입생 진로지도와 취업지원을 위해 2009년도에 전북대학교에서 실시된 신입생의 진로의식 조사 자료를 이용하여, 신입생의 선호직업과 진로성숙도의 관계를 통계패키지 R을 활용하여 행렬도로 시각화하였다.

진로성숙(career maturity)이라는 개념은 진로 상담을 위하여 개발된 도구로 청소년의 직업 흥미의 발달을 연구했던 Carter (1940)와 Super (1942)의 연구를 시작으로 1950년대 직업 선택에 관한 발달 이론가 Super (1951)와 Crites (1961)에 의해 발전되었다. 진로성숙에 대한 정의는 학자마다 조금씩 다르게 사용되어 왔는데 활발한 연구로 인하여 어느 정도 틀이 잡혀 있다. 국내에서는 김현옥과 김충기 (1989)은 진로의식, 탐색 및 의사결정에서의 발전적 변화 개념으로서 개인이 진로를 탐색하고 선택할 때 자아와 환경을 인식하여 진로를 선택하고 적응해나가는 준비의 정도라고 정의하였다. 조남태 (1990)는 개인이 진로를 선택하거나 의사결정을 할 때 자신을 정확히 평가하고 현실과 타협하여 현명하게 진로를 선택해가는 준비의 정도라고 정의하였다. 공인규 (2008)는 자신의 성격, 활동, 유능감, 직업적성, 능력

⁴ 교신저자: (561-756) 전북 전주시 덕진구 덕진동 1가 664-14, 전북대학교 통계학과(응용통계연구소), 교수.
E-mail: sschung@jbnu.ac.kr

평정의 이해를 기초로 자신의 진로를 합리적으로 계획하고 선택하는 태도와 능력이라고 정의하였다. 본 연구는 진로성숙을 개인이 진로를 선택할 때, 주위 환경과 현실에 맞추어 적절한 진로 선택을 준비하는 정도라고 정의하고, 이를 측정하기 위해 개발된 진로성숙도검사(CMI, Crites, 1973)의 태도척도 가운데 상담용 척도(counselling form B-1)를 바탕으로 작성된 설문문항 분석을 통하여 이루어졌다.

본 논문은 총 4절로 구성되어 있다. 2절에서는 행렬도의 방법론적 설명과 분석에 사용된 소프트웨어의 사용법에 대해 설명하였다. 3절은 분석 자료에 대한 서술과 행렬도를 이용한 자료의 시각화와 결과 해석에 할애하였다. 마지막으로 4절에서 본 연구를 맺는다.

2. 행렬도

2.1. 방법론

자료를 시각적으로 표현하는 방법인 행렬도는 자료행렬을 계수(rank)가 2 또는 3인 행렬로 근사시키는 행렬 이론을 이용하여 Gabriel (1971)에 의해서 제안되었는데, 행렬도에 대한 전반적인 이론적 배경에 대해선 최용석 (1992)에 자세히 소개되어 있다. 이후 Gower와 Hand (1996)는 Gabriel (1971)의 전통적인 행렬도를 산점도의 확장된 개념으로 발전시켰다. 이변량 자료의 경우 산점도를 통하여 관측치와 변수에 대한 정보를 동시에 제공할 수 있지만, 다변량 자료의 경우 산점도를 통한 자료의 시각화에는 한계가 있어 이를 극복하기 위해 제안된 방법이다. 이들은 보통의 산점도와 유사하게 여러 변수들을 본래 측정된 척도로 눈금이 매겨진 축으로 표현하였다. 두 개 이상의 변수를 표현하기 위해서 이 축들은 보통의 산점도와 같이 수직적이지는 않지만 여전히 유사한 방식으로 이용될 수 있다. 행렬도는 점과 축의 결정 방법에 따라 여러 종류로 나눌 수 있는데, 본 논문에서는 자료 분석을 위해 주성분분석(Principal Component Analysis)에 의해 결정되는 PCA 행렬도와 정준변량분석(Canonical Variate Analysis)에 의해 결정되는 CVA 행렬도, 두 종류의 행렬도를 고려하였다. 주성분분석은 작은 개수의 서로 상관되어 있지 않은 변수, 즉 주성분을 도출하여 다변량 자료가 가지고 있는 변이를 잘 설명하고자 하는 반면, 정준변량분석은 서로 다른 집단에서 관측된 다변량 자료의 경우 집단 구조를 분석하기 위한 방법으로써 집단내 분산(within-groups variance)에 대한 집단간 분산(between-groups variance)의 비를 최대로 하는 변수들의 선형 결합을 찾는다. 정준판별분석(canonical discriminant analysis)이라고도 종종 불린다. 즉 집단변수가 있는 경우 집단이 구분되는 경향을 파악하기 위해서는 CVA 행렬도가 바람직할 것이다.

PCA 행렬도는 이차원 상에 관측치들의 다차원 변동을 최적으로 표현하는데 목적이 있다. 자료의 i 번째 관측치 \mathbf{x}_i 는 이 행렬도에서 표본 공분산 행렬 또는 상관행렬의 가장 큰 두 개의 고유치에 대응하는 정규화된 고유벡터 e_1, e_2 로 정의 되는 제일과 제이 주성분 점수 $((\mathbf{x}_i - \bar{\mathbf{x}})'e_1, (\mathbf{x}_i - \bar{\mathbf{x}})'e_2)$ 를 좌표로 하는 점으로 나타내고 있다. j 번째 변수에 해당하는 축은 위의 두 정규화된 고유벡터 각각의 j 번째 성분 (e_{1j}, e_{2j}) 으로 그 방향이 결정되고, 눈금은 원점으로부터 시작하여 $w = 0, \pm 1, \pm 2, \dots$ 에 대해 위치 $w(e_{1j}, e_{2j})$ 에 그려진다. 같은 단위의 변수의 경우 축들의 눈금으로 각 변수의 산포도를 비교할 수 있다. 즉, 눈금 간격이 조밀한 축에 해당하는 변수는 다른 변수에 비해 산포도가 큰 것을 알 수 있다. 또한, 축들이 이루는 각과 눈금의 증가 방향으로 변수간의 상관관계도 알 수 있다. 고차원 자료를 이차원에 투영하여 그려진 행렬도에서 두 축이 이루는 각을 두 변수간의 상관관계 크기를 반영한다고 성급히 판단하기는 어려우나, 대체로 이루는 각이 작을수록, 상관관계가 큰 경우가 많다. 그리고 눈금의 증가 방향이 같다면 양의 관계가 있음을 알 수 있다. CVA 행렬도는 정준 공간(canonical space) 상에 모든 관측치들과 변수들과 함께 집단 평균을 점으로 나타내고 있으며, 점 좌표를 찾는 과정은 수학적으로 정준 공간에서 양측 고유치 문제(two-sided eigenvalue problem)를 해결하는 것이라고 볼 수 있는데, 개념적으로 복잡

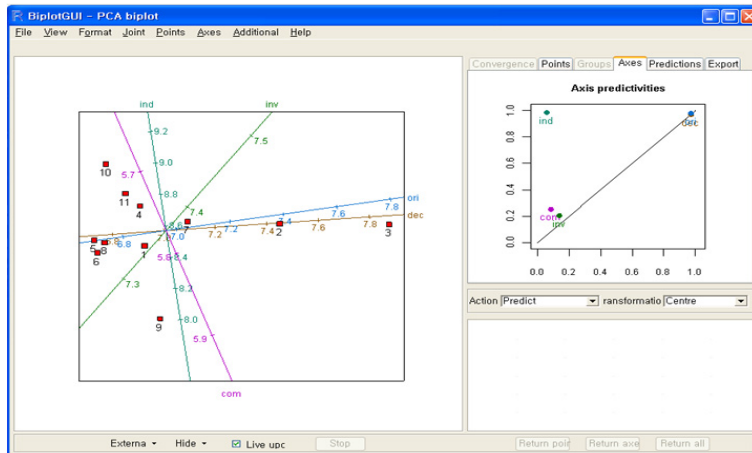


그림 2.1. BiplotGUI 사용(PCA)

하여 그 내용은 본 논문에서는 생략하고, 실제 자료 분석에서 집단이 어떻게 잘 구분되는지 확인하는데 사용하였다.

2.2. 분석도구

전통적인 행렬도는 R에서 기본적으로 제공하는 biplot() 함수를 사용하여 그려질 수 있다. 본 논문에서는 Gower와 Hand (1996)가 제안한 행렬도를 기반으로 한 R라이브러리 BiplotGUI를 사용하였는데, 양질의 시각적 결과물을 제공하여 데이터들의 관계를 사용자에게 보여주고, 그래프를 쉽게 편집할 수 있는 장점이 있다. 아래에 BiplotGUI의 간단한 사용법을 설명하고 있으나, 패키지의 다양한 기능들에 대한 자세한 설명은 la Grange 등 (2009)을 참조하면 될 것이다.

BiplotGUI를 사용하기 위해서는 먼저 다음과 같이 패키지를 다운로드 및 설치한 후 패키지를 불러온다.

```
>install.packages("BiplotGUI")
>library("BiplotGUI")
```

데이터 파일을 읽어온 후, BiplotGUI패키지의 함수인 Biplots()를 이용하여 행렬도를 얻을 수 있다.

```
>job <- read.table("파일경로 및 파일명", header=T, sep="\t")
>Biplots(job[,-1], PointLabels=job[,1])
```

이 경우 job이라는 data의 첫째 열이 이름을 나타내고 있어서, PointLabels을 통하여 행렬도 내의 각 점에 이름을 붙여 줄 수 있었다.

그림 2.1의 가장 상단의 메뉴들을 통하여 행렬도의 그래프적 표현을 조절할 수 있으며, 행렬도의 적합도, 추정치 등을 알아보기 위하여 오른쪽 상단 창 위의 탭을 이용할 수 있다. 오른쪽 하단에 위치한 창 바로 위에 Transformation이라는 메뉴가 있는데, 이를 통하여 자료의 log변환, centering, scaling 등을 할 수 있다. 또한, 행렬도 내의 각 점을 클릭하면 그 점들이 갖는 각 축(변수)에서의 값을 알 수 있다. CVA행렬도를 그리기 위해서는 먼저 다음과 같이 group으로 집단변수를 지정해 주어야 된다. 그리고 그림 2.2와 같이 메뉴 바에서 joint→CVA를 선택하면 된다.

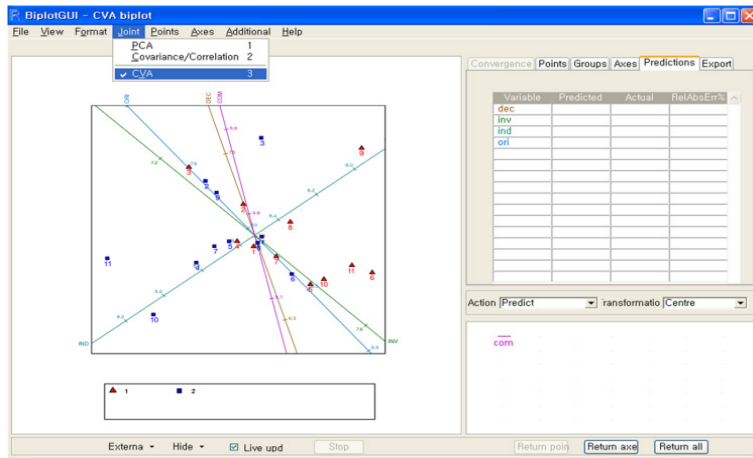


그림 2.2. BiplotGUI 사용(CVA)

표 3.1. 진로성숙도 검사 하위영역

하위 영역	내용
결정성(DEC)	선호하는 진로의 방향에 대한 확신의 정도
관여성(INV)	진로선택 과정에의 능동적 참여의 정도
독립성(IND)	진로선택을 독립적으로 할 수 있는 정도
성향성(ORI)	진로결정에 필요한 사전이해와 준비의 정도
타협성(COM)	진로선택시에 욕구와 현실을 타협하는 정도

```
>job.cva <- read.table("파일경로 및 파일명", header=T, sep="\t")
>Biplots(job.cva[,-1:-2], group=job.cva[,2], PointLabels=job.cva[,1])
```

다음 3절에서는 BiplotGUI를 이용하여 실제 자료를 분석하였다.

3. 진로성숙도 자료의 시각화

3.1. 분석 자료

전북대학교에서는 신입생의 진로지도 및 취업지원을 돕기 위한 기초자료를 수집하고, 신입생의 교육에 대한 요구사항을 파악하기 위하여 2008년도부터 신입생을 대상으로 온라인 설문조사를 실시하여 왔다. 본 연구에서는 2009년도 조사에서 얻어진 2,278명의 자료를 사용하여, 선호직업과 진로성숙도의 관계를 알아보았다.

설문 문항 중 15개 문항이 진로성숙도를 알아보기 위한 문항으로 Crites (1973)의 진로성숙도검사를 토대로 5개의 영역별(결정성(decisiveness), 관여성(involvement), 독립성(independence), 성향성(orientation), 타협성(compromise): 표 3.1 참조) 3문항씩으로 구성되어 있으며, 각 문항은 4점 척도(전혀 그렇지 않다, 그렇지 않다, 그렇다, 매우 그렇다)를 사용하여 측정하였다. 본 논문에서는 진로성숙도의 각 하위영역 3개 문항 점수를 합하여, 5개의 변수, DEC, INV, IND, ORI 그리고 COM을 생성하였다.

다음으로, 선호직업 문항인 희망 취업 분야에 대한 설명은 표 3.2와 같다. 응답자의 인구통계학적 특

표 3.2. 희망 취업분야 분류

변수	코딩	내용	코딩	내용
취업분야	1	경영·금융·회계·사무	2	교육·연구·법률·사회복지
	3	의료·보건	4	문화·예술·방송·디자인
	5	영업·판매·TM	6	서비스직
	7	건설·기계·환경·운전·운송	8	전기전자·재료·화학·섬유·식품
	9	인쇄·목재·공예·농림어업·단순직	10	IT·프로그램개발·웹
	11	기타 분야		

성을 살펴보면, 남학생이 1,309명(57.5%), 여학생이 969명(42.5%)이었고, 입학형태별로는 정시모집이 1,542명(68.0%)으로 절반 이상을 차지하였으며, 수시모집이 661명(29.1%), 기타 66명(2.9%)이었다.

3.2. 주성분 분석과 PCA행렬도

선호직업과 진로성숙도의 관계를 알아보기 위해 PCA행렬도를 통해 자료를 시각화하여 그 특징을 살펴 보았다. 이와 더불어, 주성분분석을 수행하여, 진로성숙도를 측정하는 5개 하위영역의 선형결합으로 이루어진 주성분을 찾고, 그들의 설명력을 구하여 PCA행렬도의 자료 설명정도를 알아보았다. 먼저, 희망 취업분야 11개 범주 각각에 대한 응답자들의 진로성숙도 하위영역별 평균을 구하여, 행과 열의 수가 각각 11과 5인 자료행렬 생성하였다. 그림 2.1의 오른쪽 상단 창 바로위에 위치한 Export 탭 → Vr, the matrix of basis vectors를 선택하여, 제1과 2주성분을 구하고, 각 주성분의 설명력을 나타내는 고유치는 Export 탭 → eigen을 선택하여 다음과 같은 결과를 얻었다.

$$\text{제1주성분} = 0.72 \times \text{DEC} + 0.10 \times \text{INV} - 0.12 \times \text{IND} + 0.67 \times \text{ORI} + 0.06 \times \text{COM},$$

$$\text{제2주성분} = 0.06 \times \text{DEC} + 0.14 \times \text{INV} + 0.98 \times \text{IND} + 0.11 \times \text{ORI} - 0.17 \times \text{COM}.$$

제1주성분은 결정성과 성향성에 관련된 성분으로 진로에 대한 확신과 준비 정도를 나타낸다고 볼 수 있으며, 제2주성분은 독립성에 관련된 성분으로 독립적인 진로선택 정도를 나타낸다고 볼 수 있다. 나머지 영역 관여성과 타협성은 두 주성분으로 각각 20%, 25% 정도만이 설명되고 있다. 그러나 제1주성분은 총 변량의 72%를, 제2주성분은 19%를 설명할 수 있어서, 두 개의 주성분으로 자료 변동의 91%가량이 설명되어, 주성분 두 개가 자료를 충분히 잘 요약하는 것을 알 수 있다. 그림 3.1은 2절에서 설명된 대로 R에서 자료를 불러들여 BiplotGUI패키지를 사용하여 그려진 PCA행렬도로 5차원상의 자료를 2차원상의 산점도로 나타낸 것이다.

그림 3.1에서 두 축사이의 각이 가장 작은 경우는 결정성(DEC)과 성향성(ORI)인데, 실제로 이들의 상관관계수가 0.95로 가장 큰 것을 확인 할 수 있고, 축 상에 나타난 눈금을 참조하면 결정성이 증가할 때 성향성도 증가하는 것을 볼 수 있어 이들이 양의 상관관계에 있음을 알 수 있다. 그리고 결정성(DEC)과 성향성(ORI) 축이 거의 수평이고, 독립성(IND)이 거의 수직임을 볼 수 있는데, 이는 위의 두 개의 주성분 성질과 일치함을 알 수 있다. 2번과 3번 점이 특별히 DEC와 ORI 값이 큰 것을 볼 수 있는데, 이는 교육·연구·법률·사회복지 분야와 의료·보건 분야에 취업을 원하는 학생들은 진로에 대한 확신과 준비 정도가 상당히 높다는 것을 의미한다. 그리고 4, 10, 11번 점들이 유사한 진로성숙도를 갖는 군을 형성하고 있는데, 이는 문화·예술·방송·디자인 분야, IT·프로그램개발·웹 분야, 기타 분야는 독립성이 높고 타협성이 낮아 진로선택을 현실과 타협하지 않고 독립적으로 하는 경향이 강한 분야임을 알 수 있다. 다음으로 5, 6, 8번 점들이 DEC, INV와 ORI가 모두 낮은 또 다른 군을 형성하고 있는데, 이들은 영업·판매·TM 분야, 서비스직, 전기전자·재료·화학·섬유·식품 분야에 해당하며, 선택한 진로에 대한 확신, 준

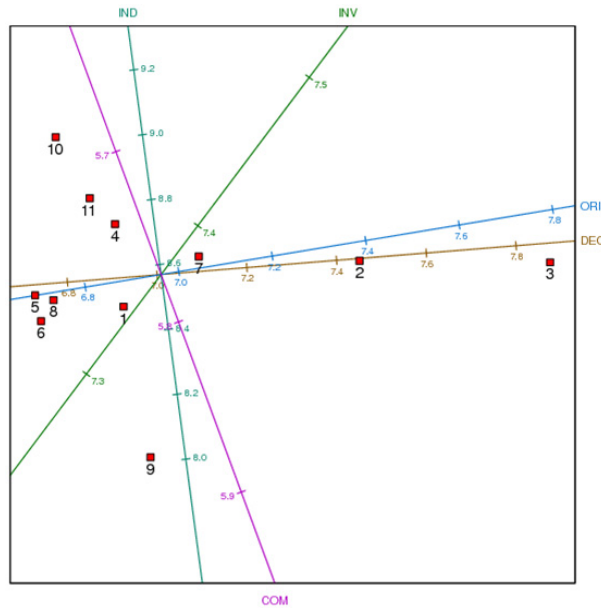


그림 3.1. PCA행렬도

표 3.3. 집단별 진로성숙도 *t*-검정 결과

특성	구분	평균±표준편차	<i>t</i> 값	유의 확률
성별	남	35.91±4.76	0.40	0.69
	여	35.98±4.44		
입학형태	수시모집	35.66±4.53	-1.94	0.05
	정시모집	36.08±4.65		

비, 참여도가 낮은 것을 볼 수 있다. 9번 점에 해당하는 인쇄·목재·공예·농림어업·단순직 분야는 진로 선택이 독립적으로 이루어지고 있지 않음을 짐작할 수 있다. 끝으로, 최근 많은 통계학과 학생들이 선호하는 경영·금융·회계·사무 분야는 모든 영역에서 거의 중간 점수를 나타내고 있어, 진로성숙도가 높지 않음을 알 수 있다.

3.3. 진로성숙도의 집단 비교와 CVA행렬도

먼저, 성별과 입학형태별로 전체적인 진로성숙도가 차이가 나는지 알아보았다. 전체적인 진로성숙 정도를 수량화하기 위하여 이들 5개의 변수들을 합하여 사용하였다. 집단별로 전체적인 진로성숙도를 *t*-검정을 통하여 비교해 보았다. 입학형태에서는 전체 2.9%를 차지하는 기타 범주를 제외한 수시모집과 정시모집만을 비교하였다. 표 3.3은 이를 정리한 결과이다. 진로성숙도는 성별차이는 크게 없어 보이나, 입학형태별로 구분하여 볼 때, 수시모집 입학생이 정시모집 입학생에 비해 진로성숙도가 낮음을 알 수 있다.

다음으로, 전체적인 진로성숙도가 차이가 나는 결과를 보인 입학형태를 집단변수로 하여 BiplotGUI패키지로 CVA 행렬도를 그려서 이들 간의 관계를 비교해 보았다. 이를 위해 입학형태별로 희망취업분야 11개 범주 각각에 대한 응답자들의 진로성숙도 하위영역별 평균을 구하여, 행과 열의 수가 각각 22와

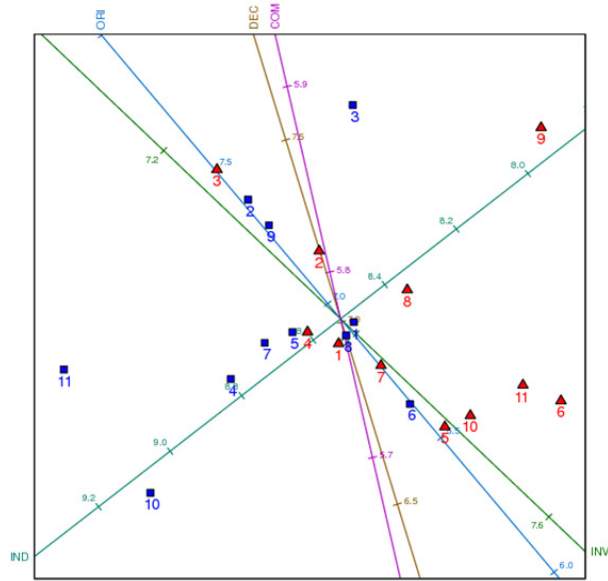


그림 3.2. CVA행렬도

5인 자료행렬 생성하였다. 그림 3.2는 수시모집(▲) 입학생과 정시모집(■) 입학생의 희망 취업분야와 진로성숙도를 비교하기 위해 그린 CVA행렬도이다. 수시모집이 정시모집에 비해 독립성과 성향성이 낮고 관여성이 높은 경향을 보이고 있다. 반면에 취업분야 범주 3번이 정시모집이 수시모집보다 독립성이 떨어짐을 볼 수 있다. 즉, 의료·보건 분야 선호하는 학생인 경우 다른 분야와 달리 수시입학생이 정시입학생보다 진로선택을 독립적으로 하였음을 짐작할 수 있다. 경영·금융·회계·사무 분야는 5개 영역 모두에서 수시와 정시 간에 큰 차이를 보이고 있지 않다.

4. 결론

다차원 자료의 시각화 방법으로 비교적 근래에 제안된 행렬도 방법은 점과 축의 결정 방법에 따라 여러 종류의 행렬도가 존재하는데, 본 논문에서는 다차원 자료의 차원축소에 많이 사용되는 주성분 분석을 이용한 PCA행렬도와 집단 판별에 사용될 수 있는 정준판별분석을 이용한 CVA행렬도를 중심으로 살펴본다. 행렬도가 탐색적 자료 분석 도구로 매우 유용함을 3절의 자료 분석 사례를 통하여 확인할 수 있었다. PCA행렬도를 통하여 각 변수와 관측치 사이의 관계를 한 눈에 볼 수 있었고, CVA행렬도는 집단별 특성을 변수와의 관계를 통하여 시각적으로 파악할 수 있게 해주었다. 행렬도의 구현을 위해 본 연구에서 사용된 무료 소프트웨어인 R의 BiplotGUI패키지의 다양하고 편리한 기능은 행렬도의 응용의 길을 더욱 넓힐 것으로 기대 된다.

참고문헌

공인규 (2008). <Holland진로탐색프로그램이 학습부진아의 진로성숙도, 진로효능감, 진로정체감 및 진로장애에 미치는 효과>, 부산대학교 대학원 교육학과 박사학위논문.
 김현옥, 김충기 (1989). 청소년의 진로성숙과 관련변인과의 상관관계, <직업교육연구>, 8, 143-160.

- 조남태 (1990). <사회계층에 따른 중학생의 진로성숙도 분석>, 숭실대학교 산업대학원 공업교육학과 석사학위논문.
- 최용석 (1992). 주성분 Biplot의 이론과 응용, <자연과학 논문집> (부산대학교), **53**, 43-49.
- Carter, H. D. (1940). Resources for the consultant: The development of vocational attitudes, *Journal of Consulting Psychology*, **4**, 185-191.
- Crites, J. O. (1961). A model for the measurement of vocational maturity, *Journal of Counselling Psychology*, **8**, 255-259.
- Crites, J. O. (1973). *Theory and Research Handbook for the Career Maturity Inventory*, CBT/McGraw-Hill, Monterey, Calif.
- la Grange, A., le Roux, N. and Gardner-Lubbe, S. (2009). BiplotGUI: Interactive biplots in R, *Journal of Statistical Software*, **30**, 1-37.
- Gabriel, K. R. (1971). The biplot graphical display of matrices with application to principal component analysis, *Biometrika*, **58**, 453-467.
- Gower, J. C. and Hand, D. J. (1996). *Biplots, Monographs on Statistics and Applied Probability*, Chapman & Hall, London.
- Super, D. E. (1942). *The Dynamics of Vocational Adjustment*, Harper & Row, New York.
- Super, D. E. (1951). The criteria of vocational success, *Occupations*, **30**, 5-8.

The Use of a Biplot in Studying the Career Maturity of College Freshmen

Hyemi Choi¹ · Chan-yong Park² · Sang-hyeop Lee³ · Sung-suk Chung⁴

¹Department of Statistics, Chonbuk National University

²Department of Statistics, Chonbuk National University

³Department of Statistics, Chonbuk National University

⁴Department of Statistics, Chonbuk National University

(Received March 2010; accepted August 2010)

Abstract

Biplot is a modern graphical methodology allowing for the projection of high-dimensional data to a low-dimensional subspace that is rich in information on variation in the data, correlation among variables as well as class separation. For the construction of biplots, we use a BiplotGUI package in a free statistical software R with increasing popularity. Moreover, using data from questionnaires given to Chonbuk National University freshmen in 2009, the relationship between career goals and career maturity are studied by applying the biplot method.

Keywords: Biplot, R package, principal component analysis, career maturity.

⁴Corresponding author: Professor, Department of Statistics (Institute of Applied Statistics), Chonbuk National University, Jeonbuk 561-756, Korea. E-mail: sschung@jbnu.ac.kr