

공변량요인 효과를 제거한 편정준상관 행렬도와 프로크러스티즈 분석을 응용한 남자 테니스선수의 체력요인 및 기초기술요인에 대한 분석연구

최태훈^a, 최용석^{1,b}

^a안동과학대학 스포츠레저과, ^b부산대학교 통계학과

요약

일반화 정준상관 행렬도(generalized canonical correlation biplot)는 정준상관분석에서 세 변수군 이상에 의해서 측정된 다변량 자료에서 변수 집단 간의 관계와 개체들의 관계를 탐색하기 위한 2차원 그림이다. 최근에 이를 활용하여 최태훈과 최용석 (2010)은 2004년 대한테니스협회(KTA)에 등록된 남자선수들 중 상위 50명을 대상으로 세 변수군인 체격요인변수군, 체력요인변수군 그리고 기초기술요인변수군의 상호 연관성을 살펴보았다. 그러나 이들 분석에서 체격요인변수군이 나머지 두 변수군과 독립적이지 못하고 선형적 영향을 미치는 것으로 판단되어 이를 공변량변수군으로 고려하였다. 이와같이 세 변수군에서 한 변수군이 공변량(covariate)으로 영향을 주는 경우 이를 제거한 정준상관분석을 편(partial)정준상관분석이라 하며 이와 관련된 편정준상관 행렬도를 염아립과 최용석 (2011)은 제안하였다. 본 연구에서는 최태훈과 최용석 (2010)의 분석에서 체격요인변수군의 영향을 제거하고 체력요인변수군과 기초기술요인변수군의 관계를 살펴보는 편정준상관 행렬도의 활용의 예를 보이고 기존 연구의 일반화 정준상관 행렬도, 편정준상관 행렬도, 정준상관 행렬도의 결과를 서로 비교하고자 한다. 덧붙여 이들 행렬도간의 형상변동 차이를 프로크러스티즈 분석을 활용하여 비교하고자 한다.

주요어: 공변량변수, 편정준상관분석, 행렬도, 프로크러스티즈 분석.

1. 서론

편정준상관 행렬도(partial canonical correlation biplot; PCCB)는 세 변수군에 대한 일반화 정준상관분석(canonical correlation analysis)에서 두 변수군에 선형적 영향을 미치는 공변량변수(covariate variables)로 이루어진 한 변수군의 영향을 제거한 후 나머지 두 변수군에 대한 편정준상관분석에서 이들 간의 관계와 개체들의 관계를 탐색하는 2차원 그림이다 (염아립과 최용석, 2011). 일반적으로 행렬도는 Gabriel (1971)에 의해서 개발되었고 국내에선 Choi (1991)가 이것의 저항버전을 개발하면서 처음으로 소개하였고, 허명희 (1993, 5장)가 국내에서 Biplot을 행렬도라 처음 불렀다. 특히, 행렬도는 복잡한 다변량 자료를 포함하여 다변량 분석의 결과를 보다 쉽게 그림으로 파악할 수 있기 때문에 최근 여러 분야에서 이에 대한 활발한 연구와 응용이 이루어지고 있다.

대개 세 변수군 이상의 경우에 적용되는 일반화 정준상관 행렬도(generalized canonical correlation biplot; GCCB)는 최용석 (2006, 1장)의 한 변수군의 행렬자료에 대한 행렬도를 세 변수군의 행렬자료로 확장한 것으로 볼 수 있다. 국내에선 Park과 Huh (1996)가 정준상관분석에서 수량화 방법(quantification method) 관점에서 2차원 그림을 제안하였고 이를 정준상관 행렬도라 하였으며, 세 변수 집단 이상인 경우까지 확장한 정준상관분석의 일반화를 시도하여 이를 허명희 (1999, 6장)에 수량화

¹ 교신저자: (609-735) 부산 금정구 장전동 산30, 부산대학교 통계학과, 교수. E-mail: yschoi@pusan.ac.kr

플롯으로 잘 요약하고 있다. 특히, 정준상관 행렬도를 GCCB와 PCCB와 구분하기 위하여 단순정준상관 행렬도(simple canonical correlation biplot; SCCB)라 하자.

이들 정준상관 행렬도가 최근에 스포츠분야에서 응용되었다. 두 변수군에 대한 SCCB의 경우 최태훈과 최용석 (2008)은 KLPGA 선수를 대상으로 기술요인변수군과 경기성적요인변수군 간의 관련성을 살폈고, 최태훈 등 (2009)은 테니스 그랜드 슬램대회 선수특성요인과 경기요인에 대한 분석을 하였다. 더 나아가 최태훈과 최용석 (2010)은 2004년 대한테니스협회(KTA)에 등록된 랭킹 100위권 이내의 선수 50명을 대상으로 세 변수군인 체격요인변수군, 체력요인변수군 그리고 기초기술요인변수군의 상호 연관성을 살펴보기 위해 GCCB를 활용하였다. 홍현욱 등 (2010)은 SCCB에서 결측값이 있는 경우 형상변동(shape variability)를 연구하였고 염아림과 최용석 (2011)은 정준상관분석에서 두 변수군 외에 이들 변수군에 선형적 영향을 미치는 공변량변수군의 영향을 제거한 후 두 변수군에 대한 PCCB를 제안하고 응용의 예를 보였다.

본 연구에서는 최태훈과 최용석 (2010)의 분석에서 나머지 두 변수군과 독립적으로 여겨졌던 체격요인변수군이 나머지 두 변수군에 영향을 미치는 것으로 판단되어 이를 공변량변수군으로 고려하였다. 따라서 이것의 영향을 제거하고 두 변수군인 체력요인변수군과 기초기술요인변수군의 관계를 살펴보는 편정준상관 행렬도 PCCB의 활용의 예를 보이고 최태훈과 최용석 (2010)에서 살펴본 GCCB와 추가적으로 PCCB 그리고 SCCB의 결과를 서로 비교하고자 한다. 덧붙여 이들 행렬도간의 형상변동 차이를 프로크러스티즈 분석을 활용하여 비교하고자 한다.

이를 위해 2절에서 공변량변수군이 있는 경우 이들의 영향력을 제거한 두 변수군에 대한 PCCB의 알고리즘을 소개하고 행렬도의 형상차이를 비교하기 위한 프로크러스티즈 분석을 소개하려 한다. 3절에서는 PCCB에 대한 실제 분석의 예를 보며 기존의 GCCB와 SCCB 서로간의 결과차이를 비교하려 한다. 4절에서는 세 가지 행렬도간의 형상변동 차이를 프로크러스티즈 분석을 활용해 비교하고자 한다.

2. 편정준상관 행렬도와 프로크러스티즈 분석

2.1. 편정준상관 행렬도

최태훈과 최용석 (2010)은 허명희 (1999, 6장)를 참고로 여러 변수군 사이의 관계를 분석하는 다변량 분석기법인 일반화 정준상관분석과 관련된 일반화 정준상관 행렬도 GCCB의 대수적 알고리즘을 세 변수군인 체격요인변수군, 체력요인변수군과 기초기술요인변수군으로 한정하여 요약하여 그 활용의 예를 보여주었다. 그러나 그들의 분석에서 나머지 두 변수군과 독립적으로 여겨졌던 체격요인변수군이 나머지 두 변수군에 영향을 미치는 것으로 판단되어 이를 공변량변수군으로 고려하고자 한다. 이럴 경우 이것의 영향을 제거하고 두 변수군의 관계를 살펴보는 염아림과 최용석 (2011)의 편정준상관 행렬도 PCCB를 응용할 수 있다.

여기서 PCCB의 대수적 알고리즘을 살펴보자. p 개의 변수와 q 개의 변수로 이루어진 두 변수군 $\mathbf{x} = (x_1, \dots, x_p)'$ 와 $\mathbf{y} = (y_1, \dots, y_q)'$ 에 선형적 영향을 미치는 m 개의 공변량변수로 이루어진 공변량변수군 $\mathbf{w} = (w_1, \dots, w_m)'$ 가 있다고 하자. PCCB의 대수적 알고리즘은 공변량변수군 \mathbf{w} 의 영향을 제거한 두 변수군 \mathbf{x} 와 \mathbf{y} 에 대한 편정준상관분석의 것을 따른다. 다음으로 세 변수군 \mathbf{x} , \mathbf{y} 그리고 \mathbf{w} 는 각각 평균벡터로 $\mu_{\mathbf{x}} = (\mu_{x_1}, \dots, \mu_{x_p})'$, $\mu_{\mathbf{y}} = (\mu_{y_1}, \dots, \mu_{y_q})'$ 그리고 $\mu_{\mathbf{w}} = (\mu_{w_1}, \dots, \mu_{w_m})'$ 를 가지며 공분산행렬 Σ_{xx} , Σ_{yy} , Σ_{ww} , $\Sigma_{xy} = \Sigma'_{yx}$, $\Sigma_{xw} = \Sigma'_{wx}$, $\Sigma_{yw} = \Sigma'_{wy}$ 을 가지는 확률벡터이다. 이들에 의해 측정된 n 명의 자료에서 표본공분산행렬은 식 (2.1)과 같다.

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{xx} & \mathbf{S}_{xy} & \mathbf{S}_{xw} \\ \mathbf{S}_{yx} & \mathbf{S}_{yy} & \mathbf{S}_{yw} \\ \mathbf{S}_{wx} & \mathbf{S}_{wy} & \mathbf{S}_{ww} \end{pmatrix}. \quad (2.1)$$

이 경우 한 변수군 $\mathbf{w} = (w_1, \dots, w_m)'$ 가 공변량변수군이므로 세 변수군에 대한 일반화 정준상관분석을 실시하는 것보다 식 (2.1)로부터 공변량변수군 \mathbf{w} 의 효과를 제거한 식 (2.2)의 조건부 표본공분산행렬 $\mathbf{S}_{\cdot\cdot\cdot\mathbf{w}}$ 을 이용하여 정준상관분석을 하는 것이 바람직하다.

$$\begin{aligned} \mathbf{S}_{\cdot\cdot\cdot\mathbf{w}} &= \begin{pmatrix} \mathbf{S}_{xx\cdot\mathbf{w}} & \mathbf{S}_{xy\cdot\mathbf{w}} \\ \mathbf{S}_{yx\cdot\mathbf{w}} & \mathbf{S}_{yy\cdot\mathbf{w}} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{S}_{xx} - \mathbf{S}_{xw}\mathbf{S}_{ww}^{-1}\mathbf{S}_{wx} & \mathbf{S}_{xy} - \mathbf{S}_{xw}\mathbf{S}_{ww}^{-1}\mathbf{S}_{wy} \\ \mathbf{S}_{yx} - \mathbf{S}_{yw}\mathbf{S}_{ww}^{-1}\mathbf{S}_{wx} & \mathbf{S}_{yy} - \mathbf{S}_{yw}\mathbf{S}_{ww}^{-1}\mathbf{S}_{wy} \end{pmatrix}. \end{aligned} \quad (2.2)$$

이를 Timm (2002, Section 8.7)은 편정준상관분석이라 하며 이를 위한 알고리즘은 정준상관분석과 유사하며 다음의 두 고유체계로 이루어진다.

$$\left(\mathbf{S}_{xx\cdot\mathbf{w}}^{-1}\mathbf{S}_{xy\cdot\mathbf{w}}\mathbf{S}_{yy\cdot\mathbf{w}}^{-1}\mathbf{S}_{yx\cdot\mathbf{w}} - \lambda^*\mathbf{I}\right)\mathbf{u}^* = \mathbf{0}, \quad \left(\mathbf{S}_{yy\cdot\mathbf{w}}^{-1}\mathbf{S}_{yx\cdot\mathbf{w}}\mathbf{S}_{xx\cdot\mathbf{w}}^{-1}\mathbf{S}_{xy\cdot\mathbf{w}} - \lambda^*\mathbf{I}\right)\mathbf{v}^* = \mathbf{0}. \quad (2.3)$$

식 (2.3)에서 고유값 λ^* 는 공통으로 편정준상관의 제곱이다. 따라서 고유체계의 성질에 따라 k 번째 λ_k^* 에 대응하는 고유벡터 \mathbf{u}_k^* 와 \mathbf{v}_k^* 를 얻게 된다. 이들 고유벡터를 편정준계수벡터라 한다. 이들은 비정칙값분해

$$\mathbf{S}_{xx\cdot\mathbf{w}}^{-\frac{1}{2}}\mathbf{S}_{xy\cdot\mathbf{w}}\mathbf{S}_{yy\cdot\mathbf{w}}^{-\frac{1}{2}} = \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime} \quad (2.4)$$

를 이용하면 대수적으로 쉽게 제공된다. 여기서 크기가 $p \times r$ 과 $q \times r$ 행렬 $\mathbf{U}^* = (\mathbf{u}_1^*, \dots, \mathbf{u}_r^*)$ 와 $\mathbf{V}^* = (\mathbf{v}_1^*, \dots, \mathbf{v}_r^*)$ 은 편정준계수벡터의 직교행렬이며 대각행렬 $\mathbf{D}^* = \text{diag}(\sqrt{\lambda_1^*}, \dots, \sqrt{\lambda_r^*})$ 는 $\sqrt{\lambda_1^*} \geq \dots \geq \sqrt{\lambda_r^*} > 0$ 관계를 갖는 비정칙값이 편정준상관에 해당하고 이를 대각원소로 하고 있다.

이제 두 변수군 $\mathbf{x} = (x_1, \dots, x_p)'$ 와 $\mathbf{y} = (y_1, \dots, y_q)'$ 에 대하여 측정된 n 명의 자료행렬을 각각 크기가 $n \times p$ 와 $n \times q$ 인 \mathbf{X} 와 \mathbf{Y} 라 하고 이들은 중심화 되어 있다고 하자. 그러면 식 (2.4)로부터 i 번째 표준 편정준상관계수벡터는 각각 다음과 같다.

$$\mathbf{a}_i^* = \mathbf{S}_{xx\cdot\mathbf{w}}^{-\frac{1}{2}}\mathbf{u}_i^*, \quad \mathbf{b}_i^* = \mathbf{S}_{yy\cdot\mathbf{w}}^{-\frac{1}{2}}\mathbf{v}_i^*, \quad i = 1, \dots, r.$$

이들로 부터 구성된 표준 편정준상관계수행렬은 $\mathbf{A}^* = (\mathbf{a}_1^*, \dots, \mathbf{a}_r^*)$ 이고 $\mathbf{B}^* = (\mathbf{b}_1^*, \dots, \mathbf{b}_r^*)$ 가 된다. 따라서 이들에 의해서 공변량변수군 \mathbf{w} 의 효과를 제거한 편정준상관 행렬도 PCCB를 제안하기로 하자. 먼저 변수군 $\mathbf{x} = (x_1, \dots, x_p)'$ 에 의한 $n \times p$ 자료행렬 \mathbf{X} 에 대한 행좌표 행렬과 열좌표 행렬은

$$\mathbf{R}_{\mathbf{X}\cdot\mathbf{w}} = \mathbf{X}\mathbf{A}^*\mathbf{D}^*, \quad \mathbf{C}_{\mathbf{X}\cdot\mathbf{w}} = \mathbf{A}^*\mathbf{D}^* \quad (2.5)$$

이며 변수군 $\mathbf{y} = (y_1, \dots, y_q)'$ 에 의한 $n \times q$ 자료행렬 \mathbf{Y} 에 대한 행좌표 행렬과 열좌표 행렬은

$$\mathbf{R}_{\mathbf{Y}\cdot\mathbf{w}} = \mathbf{Y}\mathbf{B}^*\mathbf{D}^*, \quad \mathbf{C}_{\mathbf{Y}\cdot\mathbf{w}} = \mathbf{B}^*\mathbf{D}^* \quad (2.6)$$

이다.

s 차원의 편정준상관 행렬도는 식 (2.5)와 (2.6)의 행렬에서 처음 s 열을 고려한 부행렬로 이루어지며 이 s 차원의 편정준상관 행렬도의 근사도는 다음과 같다.

$$\frac{\sum_{k=1}^s \lambda_k^*}{\sum_{k=1}^r \lambda_k^*} \times 100\%.$$

덧붙여 PCCB는 다소 번거롭지만 두 변수군 \mathbf{X} 와 \mathbf{Y} 를 공변량변수군 \mathbf{W} 를 독립변수로 하여 각각 회귀 분석한 후 얻어지는 잔차행렬(residual matrix) $\mathbf{E}_{\mathbf{X}}$ 와 $\mathbf{E}_{\mathbf{Y}}$ 를 새로운 두 변수군으로 하여 단순정준상관 행렬도 SCCB를 구하는 것과 동일한 결과를 얻을 수 있다.

2.2. 프로크루스티즈 분석

프로크루스티즈 분석(procrustes analysis)은 기하적 공간상에서 형상점(landmarks)에 의해서 나타난 개체들의 형상을 측정하고, 기술하며 비교하는 형상분석(shape analysis)에서 개체간의 형상비교를 하기 위해 한 개체를 다른 개체 쪽으로 적합시키는 방법이다 (최용석과 현기홍, 2006; Choi 등, 2005a, 2005b; 최태훈 등, 2009). 이 분석에 대한 알고리즘을 최태훈 등 (2009)의 것을 재인용하기로 하자.

X_r 과 X_s 를 개체간의 형상을 나타내는 크기가 $k \times m$ 인 중심화 형상좌표행렬이라 하면 벡터 t 와 직교행렬 R 에 의해서 X_r 의 l 번째 점 $x_{r(l)}$ 의 좌표점을 변환한 $Rx_{r(l)} + t$ 를 고려하자. 그러면 X_s 의 l 번째 점 $x_{s(l)}$ 과 $Rx_{r(l)} + t$ 간의 제곱거리합

$$\sum_{l=1}^k (x_{s(l)} - Rx_{r(l)} - t)^T (x_{s(l)} - Rx_{r(l)} - t) \quad (2.7)$$

을 생각할 수 있다. 식 (2.7)을 최소화 하는 R 과 t 를 찾는 것이 두 형상 X_r 과 X_s 가 잘 일치되도록 하게 하는 정보를 제공한다. 실제로 비정칙치분해(singular value decomposition) $X_r^T X_s = U\Lambda V^T$ 를 이용하여 $\hat{R} = VU^T$ 과 $\hat{t} = 0$ 을 제공받을 수 있고 이를 이용하여 두 형상의 일치성을 평가하는 측도(measure)인 프로크루스티즈 통계량

$$PS(X_r, X_s) = \text{tr}(X_r^T X_s) + \text{tr}(X_s^T X_r) - 2\text{tr}(\Lambda) \quad (2.8)$$

을 얻을 수 있다. 식 (2.8)의 프로크루스티즈 통계량 값이 0이면 두 형상이 일치한다고 평가할 수 있다.

3. 공변량요인을 고려한 테니스 선수의 체력요인, 기초기술요인 분석

3.1. GCCB, SCCB 그리고 PCCB의 비교

2절에서 요약한 PCCB를 응용하기 위한 자료는 2003년 3월부터 2004년 8월 15일전까지 전국대회 단체전 4강 이상 입상한 대학, 실업팀에 소속된 시니어 남자 테니스 선수들 중 2004년 대한테니스협회(KTA)에 등록된 랭킹 100위권 이내의 선수 50명에 대한 8개의 체격요인변수군(신장, 체중, 상완장, 전완장, 수장, 상지장/신장 비율, 전완장/상지장 비율, 피하지방합)과 8개의 체력요인변수군(약력, 윗몸일으키기, 하버드스텝, 윗몸앞으로굽히기, 50m달리기, 제자리높이뛰기, 눈감고외발서기, 전신반응시간), 그리고 기초기술요인변수군(그라운드 스트로크, 서비스, 발리)인 세 변수군에 의해 측정된 것이다.

이미 최태훈과 최용석 (2010)은 GCCB를 통해 시각적으로 이들을 동시에 비교하였다. 특히, 본 연구에서는 체격요인변수군이 체력요인변수군과 기초기술요인변수군에 영향을 끼치는 공변량변수(covariate variables)군으로 여겨져 GCCB분석 보다는 이들의 선형적 영향력을 제거한 편정준상관분석을 적용한 PCCB를 통한 체력요인변수군과 기초기술요인변수군의 관계를 규명하려한다. 덧붙여, GCCB와 체격요인변수군을 고려하지 않은 SCCB와의 비교를 통해 그 차이점을 설명하고 세 행렬도의 형상차이를 프로크루스티즈분석을 활용하여 비교하고자 한다.

먼저 전체 선수자료에 대하여 그림 1과 그림 2의 (a)는 각각 체격요인변수군, 체력요인변수군과 기초기술요인변수군으로 이루어진 세 변수군에 대한 2차원 GCCB 중 체력요인변수군과 기초기술요인변수군의 GCCB이고, (b)는 각각 오직 체력요인변수군과 기초기술요인변수군으로 이루어진 두 변수군에 대한 2차원 SCCB이고, (c)는 각각 공변량변수군인 체격요인변수군의 영향을 제거한 체력요인변수군과 기초기술요인변수군에 대한 2차원 PCCB이다. (d)는 PCCB에서 기초기술요인변수군에 대한 전 체선수의 행렬도이나 GCCB나 SCCB에서도 대동소이한 결과를 얻을 수 있다. 그림 1과 그림 2의 (a)

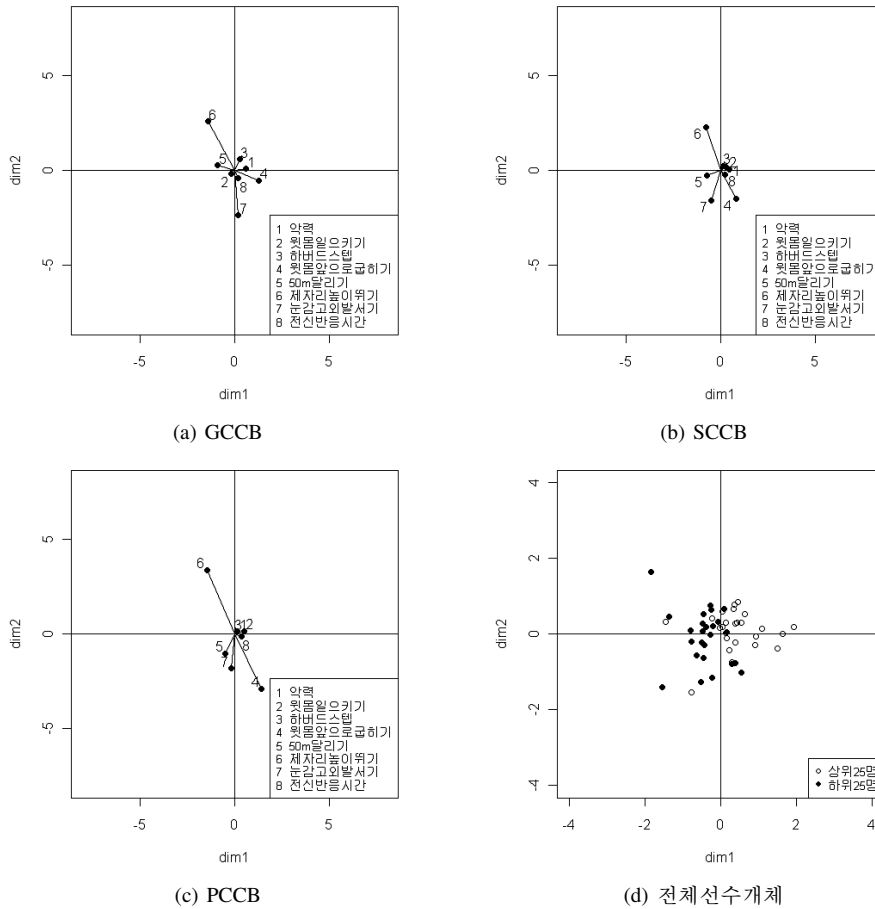


그림 1: 체력요인변수군의 GCCB, SCCB, PCCB의 비교

GCCB는 제1과 제2 일반화 정준상관이 0.65와 0.47로 근사도에서 제1과 제2 정준축이 54%와 29%로 전체 적합도는 83%이며, (b) SCCB는 제1과 제2 단순정준상관이 0.80과 0.48로 근사도에서 제1과 제2 정준축이 66%와 24%로 전체 적합도는 90%이다. (c) PCCB와 (d) PCCB에서 전체선수개체에 대해서 제1과 제2 편정준상관이 0.75와 0.63으로 근사도에서 제1과 제2 정준축이 54%와 38%로 전체 적합도는 92%이다. 전체적으로 2차원 행렬도의 적합도는 80% 이상으로 원자료를 잘 설명한다고 여겨진다.

그림 1 체력요인변수에 대한 세 행렬도 (a) GCCB, (b) SCCB와 (c) PCCB를 비교하자면 모두가 다른 경향을 보인다. 특히, 50m달리기(5)와 눈감고외발서기(7) 두 변수는 체력요인변수군을 공변량요인도 독립된 변수군으로 고려하여 세 변수군을 모두 사용한 (a) GCCB에서는 제1 정준축(dim1)에 대하여 반대쪽에 놓여 있어 서로 상관이 없는 변수임을 보여주고 있다. 이와 반대로 공변량요인을 제거한 (c) PCCB와 처음부터 공변량요인을 고려하지 않은 (b) SCCB에서는 제1 정준축(dim1)의 왼편에 두 변수가 있어 서로 상관이 높은 변수임을 보여주고 있다. 다만 PCCB에서 SCCB에 비해 상대적으로 두 변수는 상관이 매우 높은 변수임을 알 수 있다. 이는 체격변수요인을 공변량요인으로 고려하지 않고 상호 독립된 변수군으로 대등하게 고려한 GCCB와 처음부터 공변량요인을 무시한 SCCB의 문제점으로 여겨진다. 전체선수개체에 대한 행렬도 (d)에서는 제1 정준축(dim1)을 중심으로 오른편에 상위랭킹 선

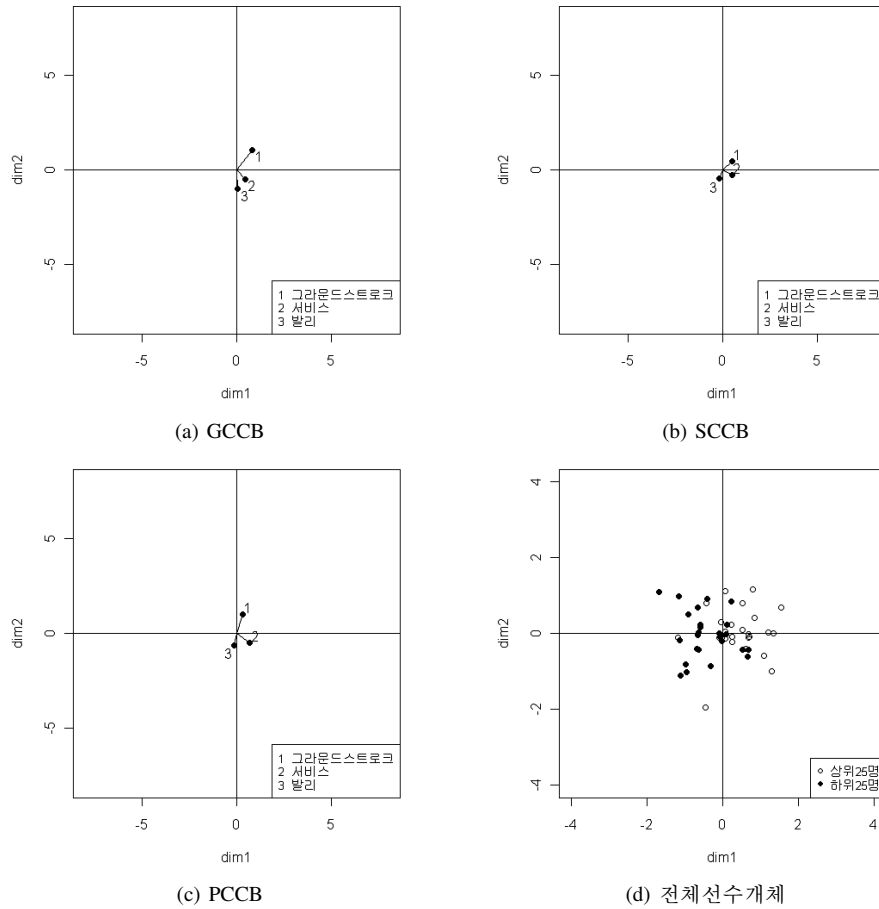


그림 2: 기초기술요인변수군의 GCCB, SCCB, PCCB의 비교

수들이 포진해 있고 왼편이 하위랭킹 선수들이 포진해 있다. 이러한 경향은 그림 2 기술요인변수군의 행렬도에 대해서도 같음을 알 수 있다.

그림 2 기초기술요인변수에 대한 세 행렬도 (a) GCCB, (b) SCCB와 (c) PCCB를 비교하자면 그림 1처럼 세 변수군을 모두 사용한 GCCB (a)만 변수들의 경향이 다소 나머지 (b)와 (c)와 달라 보인다. 특히, GCCB (a)에서 세 변수 모두가 제1 정준축(dim1) 오른편에 놓여 있어 발리(3)가 왼편에 놓여진 나머지 행렬도와는 다소 다른 경향을 보여주고 있다. 이 또한 체격요인변수군을 공변량요인으로 고려하지 않은 것에서 기인했다고 여겨진다. 그러나 발리(3)의 경우 세 행렬도에서 거의 원점 가까이 있어 상위와 하위 선수들의 기량을 가늠하는데 미치는 영향은 미미함을 알 수 있다. 따라서 기초기술요인변수군의 경우 세 행렬도는 공변량효과를 고려한다고 해도 비슷한 경향을 보여준다고 여겨진다.

3.2. 세 행렬도의 형상비교를 위한 프로크러스티즈 분석

다음으로 표 1은 3.1절에서 살펴 본 전체자료에 대한 체격요인변수군과 기초기술요인변수군별 GCCB, SCCB 그리고 PCCB의 형상을 비교하기 위한 프로크러스티즈 통계량을 제공하고 있다. 먼저 체격요인변수군에서 GCCB와 PCCB가 통계량 9.95로 제일 변동(variability)이 많고 다음으로 SCCB와

표 1: 세 행렬도의 형상비교를 위한 프로크루스티즈 통계량

	체력요인			기초기술요인		
	GCCB	SCCB	PCCB	GCCB	SCCB	PCCB
GCCB	0.00			0.00		
SCCB	3.91	0.00		1.03	0.00	
PCCB	9.95	4.52	0.00	0.59	0.30	0.00

PCCB가 4.52로 변동이 높다. 이는 그림 1에서 세 행렬도의 비교에서 지적한 것과 같이 체격변수요인을 공변량요인으로 고려하지 않고 상호 독립된 변수군으로 대등하게 고려한 GCCB와 처음부터 아예 고려하지 않은 SCCB의 문제점으로 여겨진다. 기초기술요인변수군에서는 세 행렬도의 프로크루스티즈 통계량이 작아 상호간의 변동이 높지 않음을 보여준다. 이는 그림 2에서 공변량요인으로 인하여 세 행렬도의 차이는 보여주지만 그 영향력이 다소 미미했음을 지적한 것과 맥락을 같이 한다고 여겨진다.

4. 결론

본 연구에서는 2004년 대한테니스협회(KTA)에 등록된 랭킹 100위권 이내의 남자선수 50명에 대한 최태훈과 최용석 (2010)의 일반화 정준상관 행렬도(GCCB)를 활용한 분석에서 체격요인변수군이 나머지 두 변수군(체력요인변수군과 기술요인변수군)에 영향을 미치는 것으로 판단되어 이를 공변량변수군으로 고려하였다. 따라서 이것의 영향을 제거하고 두 변수군인 체력요인변수군과 기초기술요인변수군의 관계를 살펴보는 편정준상관 행렬도 PCCB의 활용의 예를 보이고 최태훈과 최용석 (2010)에서 살펴 본 GCCB와 추가적으로 PCCB 그리고 SCCB의 결과를 서로 비교하였다. 특히, 체력요인변수군에 대한 PCCB에서 50m달리기(5)와 눈감고외발서기(7) 두 변수는 GCCB에서는 서로 상관이 없는 변수로 SCCB에서는 상관은 있으나 낮게 나타남을 보여 주었다. 이러한 행렬도간의 형상변동 차이를 프로크루스티즈 분석을 활용하여 비교하였다.

참고 문헌

염아림, 최용석 (2011). 편정준상관 행렬도, <응용통계연구>, **24**, 559-566.
 최용석 (2006). <행렬도 분석>, 기초과학 총서 2권, 부산대학교 기초과학연구원.
 최용석, 현기홍 (2006). <통계적 형상분석의 이해와 응용-프로크루스티즈 분석의 저항성 버전 연구 및 개발->, 자유아카데미, 서울.
 최태훈, 최용석 (2008). 정준상관 행렬도와 군집분석을 응용한 KLPGA 선수의 기술과 경기성적요인에 대한 연관성분석, <응용통계연구>, **21**, 429-439.
 최태훈, 최용석 (2010). 일반화 정준상관 행렬도와 프로크루스티즈 분석을 응용한 대한테니스협회 등록 선수의 체격요인, 체력요인 및 기초기술요인에 대한 분석연구, <한국통계학회논문집>, **17**, 917-925.
 최태훈, 최용석, 신상민 (2009). 테니스그랜드슬램대회의 선수특성요인과 경기요인에 대한 분석연구-정준상관 행렬도와 프로크루스티즈 분석의 응용-, <응용통계연구>, **22**, 855-864.
 허명희 (1993). <統計相談의 이해>, 자유아카데미, 서울.
 허명희 (1999). <다변량수량화>, 자유아카데미, 서울.
 홍현욱, 최용석, 신상민, 강창완 (2010). 결측값이 있는 정준상관행렬도의 형상변동 연구, <응용통계연구>, **23**, 955-966.
 Choi, Y. S. (1991). *Resistant Principal Component Analysis, Biplot and Correspondence Analysis*, Unpublished Ph.D. Dissertation, Department of Statistics, Korea University.
 Choi, Y. S., Hyun, G. H. and Kim, J. G. (2005a). A numerical comparison of map variability in SCA Using the Procrustes Analysis, *Journal of the Korean Data Analysis Society*, **7**, 1531-1538.

- Choi, Y. S., Hyun, G. H. and Yun, W. J. (2005b). Biplots' variability based on the Procrustes analysis, *Journal of the Korean Data Analysis Society*, **7**, 1925–1933.
- Gabriel, K. R. (1971). The biplot graphics display of matrices with applications to principal component analysis, *Biometrika*, **58**, 453–467.
- Park, M. R. and Huh, M. H. (1996). Canonical correlation biplot, *The Korea Communications in Statistics*, **3**, 11–19.
- Timm, N. H. (2002). *Applied Multivariate Analysis*, Springer, New York.

2011년 6월 접수; 2011년 11월 채택

Relationship between Physical Fitness and Basic Skill Factors for KTA Players Using the Partial Canonical Correlation Biplot Removing the Linear Effect of the Set of Covariate Variables and Procrustes Analysis

Tae-Hoon Choi^a, Yong-Seok Choi^{1,b}

^aDepartment of Sport and Leisure Studies, Andong Science College

^bDepartment of Statistics, Pusan National University

Abstract

The generalized canonical correlation biplot is a 2-dimensional plot to graphically investigate the relationship between more than three sets of variables and the relationship between observations and variables. Recently, Choi and Choi (2010) investigated the relationship physique, physical fitness and basic skill factors of Korea Tennis Association(KTA) players of using this biplot; however we consider the set of covariate variables affecting the linearly on two sets of variables. In this case, if we apply the generalized canonical correlation biplot, we cannot clearly interpret the other two sets of variables due to the effect of the set of covariate variables. Moreover, Yeom and Choi (2011) provided partial canonical correlation analysis that removed the linear effect of the set of covariate variables on two sets of variables. In addition, Procrustes analysis is a useful tool for comparing shape between configurations. In this study, we will investigate the relationship between physical fitness and basic skill factors of KTA players of using a partial canonical correlation biplot and Procrustes analysis. We compare shapes and shape variabilities for the generalized, partial and simple canonical correlation biplots.

Keywords: Biplot, partial canonical correlation analysis, tennis players, Procrustes analysis.

¹ Corresponding author: Professor, Department of Statistics, Pusan National University, Busan 609-735, Korea.
E-mail: yschoi@pusan.ac.kr