

테니스 그랜드슬램대회의 선수특성요인과 경기요인에 대한 분석연구 -정준상관 행렬도와 프로크러스티즈 분석의 응용-

최태훈¹ · 최용석² · 신상민³

¹안동과학대학 체육계열, ²부산대학교 통계학과, ³부산대학교 통계학과

(2009년 4월 접수, 2009년 5월 채택)

요약

정준상관 행렬도(canonical correlation biplot)는 정준상관분석(canonical correlation analysis)에서 두 변수 집단에 의해서 측정된 다변량 자료에서 변수 집단 간의 관계와 개체들의 관계를 탐색하기 위한 2차원 그림이다. 최근에 최태훈과 최용석 (2008)는 2006년도 KLPGA 선수를 대상으로 정준상관 행렬도를 통해 기술요인변수군과 경기성적요인변수군간의 관련성을 살펴보고 군집분석을 활용하여 각 선수들의 군집을 시도하였다. 프로크러스티즈 분석(Procrustes analysis)은 두 형상(shape)의 유사성을 비교하는 데 사용되는 기법이다. 본 연구에서는 테니스 그랜드슬램대회의 선수특성요인변수군과 경기요인변수군에 대한 분석연구를 정준상관 행렬도를 적용하여 살펴보고 프로크러스티즈 분석을 통하여 행렬도 형상비교를 하였다.

주요어: 정준상관분석, 정준상관행렬도, 테니스 그랜드슬램대회, 프로크러스티즈 분석.

1. 서론

정준상관 행렬도(canonical correlation biplot)는 정준상관분석(canonical correlation analysis)에서 두 변수 집단에 의해서 측정된 다변량 자료에서 변수 집단 간의 관계와 개체들의 관계를 탐색하기 위한 2차원 그림이다. 이는 일반적으로 최용석 (2006, 1장)의 한 변수 집단의 행렬자료에 대한 일반적인 행렬도를 두 변수 집단의 행렬자료로 확장한 것으로 볼 수 있다.

국내에선 Park과 Huh (1996a, 1996b)가 정준상관분석에서 수량화 방법 (quantification method) 관점에서 2차원 그림을 제안하였고 이를 정준상관 행렬도라 하였으며, 세 변수 집단 이상인 경우까지 확장한 정준상관분석의 일반화를 시도하였다.

일반적으로 행렬도는 복잡한 다변량 분석의 결과를 보다 쉽게 파악할 수 있기 때문에 최근 여러 분야에서 이에 대한 활발한 연구와 응용이 이루어지고 있다. 행렬도는 Gabriel (1971)에 의해서 개발되었고 국내에선 Choi (1991)가 이것의 저항버전을 개발하면서 처음으로 소개하였고, 허명희 (1993, 5장)가 국내에서 Biplot을 행렬도라 처음 불렀다. 더군다나 최용석 등 (2005a, 2005b)는 다변량 분산분석 모형에서 행렬도의 활용을 제안하고 응용의 예를 보였다. 최용석 등 (2005c), Choi 등 (2005a) 그리고 Choi 등 (2005b)는 행렬도에 대한 활용의 폭을 넓혀왔다.

최근에 Choi와 Kim (2008)은 대용량 자료에서 행렬도의 해석상 문제점을 극복하는 데 군집분석을 활용하는 방법을 제시하였다. 최태훈과 최용석 (2008)은 2006년도 KLPGA 선수 중 상금 순위 상위 50명

²교신저자: (609-735) 부산시 금정구 장전동 산30, 부산대학교 통계학과, 교수. E-mail: yschoi@pusan.ac.kr

을 대상으로 정준상관 행렬도를 통해 기술요인변수군과 경기성적요인변수군 간의 관련성을 살펴보고 군 집분석을 활용하여 각 선수들의 군집을 시도하였다.

본 연구에서는 테니스 그랜드슬램 선수 중 8강에 진출한 남자 선수들을 대상으로 정준 상관분석 행렬도를 통해 특성요인(키, 몸무게, 랭킹)과 경기요인(1st serve %, Aces, Double faults, Unforced errors, Winning % on 1st Serve, Winning % on 2nd Serve, Winners (Including Service), Receiving points won, Break point conversions %, Net approaches %, Total point won)의 연관성을 살펴보려 한다. 이와 더불어 승과 패, 경기코트별(하드, 양투카, 잔디) 차이점을 살펴보려한다. 2절에서 정준상관 행렬도를 소개하고 3절에서는 실제 분석의 예를 보이고자 한다. 4절에서는 승패별과 경기코트별 행렬도의 형상 차이를 비교하기위한 프로크러스티즈 분석을 활용한 예를 보이고자 한다.

2. 정준상관 행렬도와 프로크러스티즈 분석의 활용

2.1. 정준상관 행렬도

이 절에서는 최용석 (2006, 2.3절)을 참고로 정준상관분석와 관련된 정준상관 행렬도의 대수적인 면을 간단히 요약하기로 하자.

정준상관분석은 두 변수군(sets of variables) 사이의 관계를 분석하는 다변량기법이다. 먼저 p 개의 변수와 q 개의 변수로 이루어진 두 변수군 $\mathbf{x} = (x_1, \dots, x_p)^T$ 와 $\mathbf{y} = (y_1, \dots, y_q)^T$ 는 각각 평균 $\mu_{\mathbf{x}} = (\mu_{x_1}, \dots, \mu_{x_p})^T$ 와 $\mu_{\mathbf{y}} = (\mu_{y_1}, \dots, \mu_{y_q})^T$ 를 가지며 공분산행렬 $\Sigma_{xx}, \Sigma_{yy}, \Sigma_{xy} = \Sigma_{yx}$ 을 가지는 확률 벡터이다.

다음으로 임의의 계수벡터 \mathbf{u} 와 \mathbf{v} 에 대해 두 변수군 각각의 선형결합

$$\mathbf{Z}_x = u_1x_1 + \dots + u_px_p = \mathbf{u}^T \mathbf{x}, \quad \mathbf{Z}_y = v_1y_1 + \dots + v_qy_q = \mathbf{v}^T \mathbf{y} \tag{2.1}$$

을 생각하자. 따라서 식 (2.1)의 두 선형결합 \mathbf{Z}_x 와 \mathbf{Z}_y 의 상관은

$$r_{\mathbf{Z}_x \mathbf{Z}_y} = \frac{\sum_{i=1}^n z_{x_i} z_{y_i}}{\sqrt{\sum_{i=1}^n z_{x_i}^2} \sqrt{\sum_{i=1}^n z_{y_i}^2}} = \frac{\mathbf{u}^T \mathbf{S}_{xy} \mathbf{v}}{\sqrt{\mathbf{u}^T \mathbf{S}_{xx} \mathbf{u}} \sqrt{\mathbf{v}^T \mathbf{S}_{yy} \mathbf{v}}} \tag{2.2}$$

이다. 이젠 식 (2.1)에서 계수벡터 \mathbf{u} 와 \mathbf{v} 는 식 (2.2)의 두 선형결합의 상관을 최대화하는 알고리즘을 통하여 구할 수 있다. 상관을 최대화하는 알고리즘은 \mathbf{Z}_x 와 \mathbf{Z}_y 의 분산이 1인 제약조건 $\mathbf{u}^T \mathbf{S}_{xx} \mathbf{u} = 1$ 과 $\mathbf{v}^T \mathbf{S}_{yy} \mathbf{v} = 1$ 을 두고 $\mathbf{u}^T \mathbf{S}_{xy} \mathbf{v}$ 를 최대화하는 계수벡터 \mathbf{u} 와 \mathbf{v} 를 찾는 것과 동일하다. 이 알고리즘은 정준계수벡터와 정준상관을 대수적으로 한꺼번에 제공하는 비정칙값분해(singular value decomposition)

$$\mathbf{S}_{xx}^{-\frac{1}{2}} \mathbf{S}_{xy} \mathbf{S}_{yy}^{-\frac{1}{2}} \mathbf{S}_{yx} = \mathbf{U} \mathbf{D}_{\sqrt{\lambda}} \mathbf{V}^T \tag{2.3}$$

를 이용하면 편리하다. 여기서 크기가 $p \times r$ 과 $q \times r$ 행렬 $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_r)^T$ 와 $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_r)^T$ 은 정준계수벡터의 직교행렬이며 대각행렬 $\mathbf{D}_{\sqrt{\lambda}} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r})$ 는 $\sqrt{\lambda_1} \geq \dots \geq \sqrt{\lambda_r} > 0$ 의 관계를 갖는 비정칙값이 정준상관에 해당하고 이를 대각원소로 하고 있다.

특히, k 번째 정준상관 $\sqrt{\lambda_k}$ 에 대응하는 정준계수벡터 \mathbf{u}_k 와 \mathbf{v}_k 를 식 (2.1)에 대입하면 정준변수(canonical variates)의 한 짝인 $\mathbf{Z}_{xk} = \mathbf{u}_k^T \mathbf{x}$ 와 $\mathbf{Z}_{yk} = \mathbf{v}_k^T \mathbf{y}$ 를 구하게 된다.

따라서 두 변수군 $\mathbf{x} = (x_1, \dots, x_p)^T$ 와 $\mathbf{y} = (y_1, \dots, y_q)^T$ 에 대하여 측정된 n 명의 자료행렬을 각각 크기가 $n \times p$ 와 $n \times q$ 인 \mathbf{X} 와 \mathbf{Y} 라 하고 이들은 중심화되어 있다고 하자. 그러면 식 (2.3)으로부터 i 번째

표준 정준상관계수벡터는 각각

$$\mathbf{a}_i = \mathbf{S}_{xx}^{-\frac{1}{2}} \mathbf{u}_i, \quad \mathbf{b}_i = \mathbf{S}_{yy}^{-\frac{1}{2}} \mathbf{v}_i, \quad i = 1, \dots, r \quad (2.4)$$

이다. 이들에 의해서 구성된 표준 정준상관계수행렬은 $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_r)$ 와 $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_r)$ 가 된다. 이들에 의해서 계산되어진 자료행렬 \mathbf{X} 에 대한 정준상관 행렬도의 행좌표 행렬과 열좌표 행렬은

$$\mathbf{R}_X = \mathbf{XAD}\sqrt{\lambda}, \quad \mathbf{C}_X = \mathbf{AD}\sqrt{\lambda} \quad (2.5)$$

이 되며, 자료행렬 \mathbf{Y} 에 대한 정준상관 행렬도의 행좌표 행렬과 열좌표 행렬은

$$\mathbf{R}_Y = \mathbf{YBD}\sqrt{\lambda}, \quad \mathbf{C}_Y = \mathbf{BD}\sqrt{\lambda} \quad (2.6)$$

이 된다.

s 차원의 정준상관 행렬도는 식 (2.5)와 (2.6)의 행렬에서 처음 s 개 열을 고려한 부행렬로 이루어지며 이 s 차원의 정준상관 행렬도의 근사도는 일반적으로 식 (2.3)으로부터 얻어지는 전체 정준상관의 제곱값 합에서 s 개의 정준상관의 제곱값 합이 차지하는 비율을 이용한다.

2.2. 프로크러스티즈 분석

프로크러스티즈 분석(procrustes analysis)이란 기하적 공간상에서 형상점(landmarks)에 의해서 나타낸 개체들의 형상을 측정하고, 기술하며 비교하는 형상분석에서 개체간의 형상비교를 하기 위해 한 개체를 다른 개체 쪽으로 적합시키는 방법이다 (최용석과 현기홍, 2006; Choi 등, 2005a, 2005b). Choi 등 (2005b)는 이미 일반적으로 잘 알려진 주성분, 대칭 그리고 JK 행렬도의 형상비교를 위해 이 분석을 사용한 바가 있다. 따라서 이 분석에 대한 알고리즘을 간단히 설명하기로 하자.

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

$$\sum_{l=1}^k (\mathbf{x}_{s(l)}\mathbf{R}\mathbf{x}_{r(l)} - \mathbf{t})^T (\mathbf{x}_{s(l)}\mathbf{R}\mathbf{x}_{r(l)} - \mathbf{t}) \quad (2.7)$$

을 생각할 수 있다. 식 (2.7)을 최소화 하는 \mathbf{R} 과 \mathbf{t} 를 찾는 것이 두 형상 \mathbf{X}_r 과 \mathbf{X}_s 가 잘 일치되도록 하게하는 정보를 제공한다. 실제로 비정칙치분해(singular value decomposition) $\mathbf{X}_r^T \mathbf{X}_s = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$ 를 이용하여 $\hat{\mathbf{R}} = \mathbf{V}\mathbf{U}^T$ 과 $\hat{\mathbf{t}} = \mathbf{0}$ 를 제공받을 수 있고 이를 이용하여 두 형상의 일치성을 평가하는 척도(measure)인 프로크러스티즈 통계량

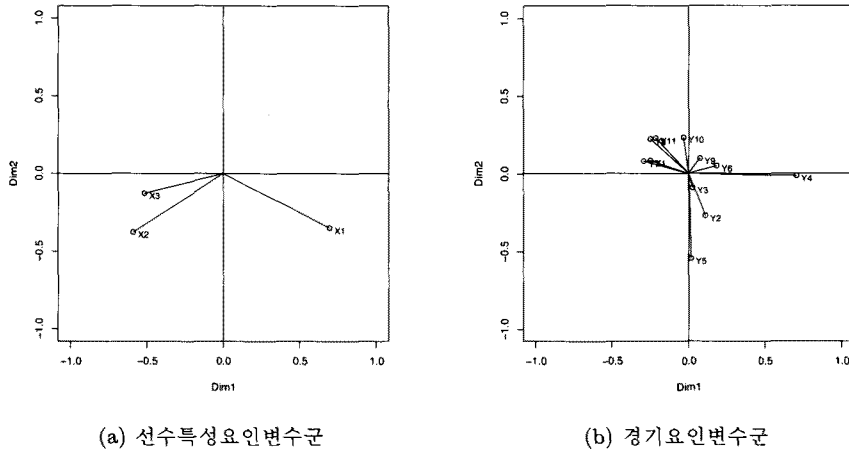
$$\text{PS}(\mathbf{X}_r, \mathbf{X}_s) = \text{tr}(\mathbf{X}_r^T \mathbf{X}_r) + \text{tr}(\mathbf{X}_s^T \mathbf{X}_s) - 2\text{tr}(\mathbf{\Lambda}) \quad (2.8)$$

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

3. 테니스 그랜드슬램 선수의 특성요인과 경기요인 분석

3.1. 전체자료 비교

2절에서 요약한 정준상관분석 행렬도를 응용하기 위한 자료는 세계 4대 테니스 그랜드슬램대회 홈페이지 (전호오픈: www.australianopen.com, 전볼오픈: www.rolandgarros.com, 전영오픈: www.wimble



(a) 선수특성요인변수군

(b) 경기요인변수군

그림 3.1. 전체자료에 대한 정준상관 행렬도

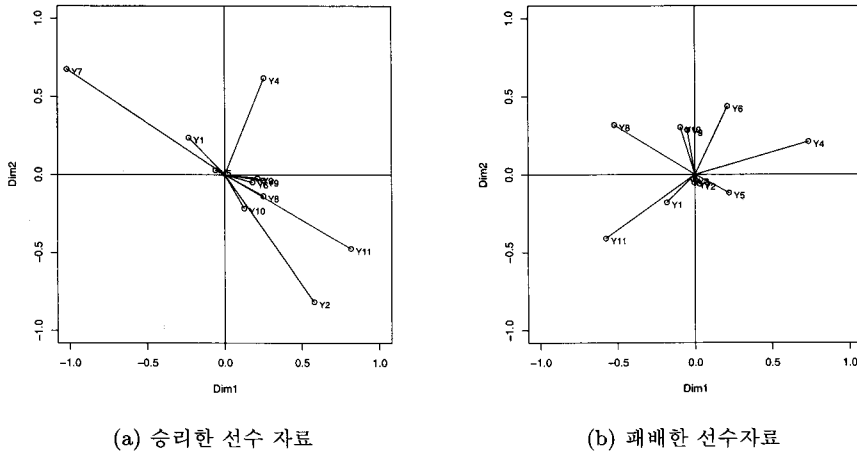
don.org, 전미오픈: www.usopen.org, 2007)에서 제공하는 2007년도 남자단식 8강부터 결승전까지 진출한 29명의 총 88경기에서 선수들의 특성요인과 경기요인을 정리한 것이다. 특히, 본 연구에서는 키(X1), 몸무게(X2), 랭킹(X3)을 특성요인변수군으로 1st serve %(Y1), Aces(Y2), Double faults(Y3), Unforced errors(Y4), Winning % on 1st Serve(Y5), Winning % on 2nd Serve(Y6), Winners Including Service(Y7), Receiving points won(Y8), Break point conversions %(Y9), Net approaches %(Y10), Total point won(Y11)을 경기요인변수군으로 고려하였다. 특히, 체중은 장시간에 걸쳐 치러는 테니스 경기에서 중요한 체격변수이며 상위 랭킹 선수들이 반드시 갖추어야 할 요소이며 키는 서비스에 아주 중요한 변수이다 (최태훈, 2005).

이미 김도연과 최태훈 (2000)에서도 테니스 그랜드슬램 대회 경기분석을 하였으나, 경기요인들을 서로 독립적으로 가정하고 남녀 간의 차이를 김정하거나 선수특성요인 중 랭킹만 가지고 경기요인간의 독립적 순위상관분석을 실시하는 등 통계적으로 다소 문제점이 었보였다.

그림 3.1(a)와 (b)는 전체 선수자료에 대한 각각 기술요인변수군과 경기성적요인변수군에 대한 2차원 정준상관 행렬도로 적합도는 77.42%이며 이 질에서 나타나는 행렬도의 적합도가 전체적으로 75%이상 93%이하로 상당히 높고 첫 번째 정준상관이 70%이상으로 수평축인 Dim1(제1축) 대한 해석만으로도 충분하다고 여겨진다.

그림 3.1(a) 선수특성요인변수군에 대한 정준상관 행렬도에서 X2(체중), X3(랭킹)이 수평축인 Dim1(제1축)에 대하여 왼쪽 방향에 놓여 있어 양의 상관이 높고 같은 경향을 나타냄을 알 수 있다. 이는 앞서 지적한 바와 같이 체중은 상위랭킹 선수들이 지녀야 할 중요한 체격변수임을 보여주고 있다. X1(신장)은 그 반대의 경향을 나타내고 있다. 선수특성요인변수군에 대한 이런 성향은 3.2절의 승패별 비교에서 승리한 선수 자료에 대한 것과 패배한 선수 자료에 대한 정준상관 행렬도와 3.3절의 경기코트별 비교에서 하드코트, 양투카코트, 잔디코트 출전 선수 자료에 대한 정준상관 행렬도에서 대동소이한 결과를 보이므로 이는 생략하기로 하고 경기요인변수군에 대한 것만 비교하기로 하겠다.

다음으로 경기요인변수군 대한 2차원 정준상관 행렬도 그림 3.1(b)에서 Y1(1st serve %), Y7(winners including service), Y8(receiving points won), Y10(net approaches %), Y11(total points won)는



(a) 승리한 선수 자료

(b) 패배한 선수자료

그림 3.2. 승패별 경기요인변수군에 대한 정준상관 행렬도

Dim1(제1축)에 대하여 왼쪽에 놓여있어 득점과 관련된 변수로 해석할 수 있으며 서비스에 관련된 나머지 변수 Y2(aces), Y3(double faults), Y4(unforced errors), Y5(winning % on 1st serve), Y6(winning % on 2nd serve), Y9(break point conversions %)와는 서로 다른 방향에 놓여 있어 성향이 다른 변수임을 알 수 있다.

덧붙여 서로 다른 변수군의 변수들 간의 연관성을 엿보기 위해선 그림 3.1(a)와 (b)를 동시에 비교해 보면 된다. 즉, Dim1(제1축)에 대하여 그림 3.1(a)의 왼편에 있는 선수특성요인변수 X2(체중)과 X3(랭킹)은 그림 3.1(b)의 득점과 관련된 경기요인변수 Y1(1st serve %), Y7(winners including service), Y8(receiving points won), Y10(net approaches %), Y11(total points won)과 같은 방향에 놓여 있어 서로 연관성이 높음을 보여준다. 특히, 체중은 장시간에 걸쳐 치르는 테니스 경기에서 중요한 체격변수이며 득점과 관련된 경기요인에서 높은 값을 나타내는 상위 랭킹 선수들이 반드시 갖추어야할 요소임을 알 수 있다. 이와는 반대로 그림 3.1(a)의 X1(신장)과 서비스에 관련된 변수인 그림 3.1(b)의 Y2, Y3, Y4, Y5, Y6, Y9가 서로 연관성이 있음을 보여준다. 이는 선수특성요인변수인 신장이 경기요인변수이면서 서비스를 수행하는 데 중요한 영향을 미치고 있음을 의미한다. 실제로 최태훈 (2005)은 2005년도 국내 대학실업 남자 테니스 선수들을 대상으로 한 연구에서 신장이 서비스에 아주 중요한 변수임을 지적하고 있다.

3.2. 승패별 비교

그랜드슬램대회에서 승리한 선수와 패배한 선수 자료에 대한 선수특성요인변수군과 경기요인변수군에 대한 정준상관 행렬도를 통한 차이를 비교하기로 하자. 그림 3.2(a) 승리한 선수 자료와 (b) 패배한 선수 자료에 대한 경기요인변수군에 대한 정준상관 행렬도이다.

승리한 선수들만의 자료에 대한 정준상관 행렬도 그림 3.2(a)에서 Dim1(제1축)에 대하여 Y1(1st serve %), Y5(winning % on 1st serve), Y7(winners including service)는 왼쪽에 놓여있어 서비스를 잘 넣어서 이길 확률과 관련된 변수로 해석할 수 있으며 이와 반대로 놓여진 나머지 변수는 Y2(aces), Y3(double faults), Y4(unforced errors), Y6(winning % on 2nd serve), Y8(receiving points won), Y10(net ap-

표 3.1. 승패별 행렬도의 형상비교를 위한 프로크러스티즈 통계량

	선수특성요인		경기요인	
	승	패	승	패
승	0		0	
패	0.101	0	4.368	0

proaches %), Y11(total points won), Y9(break point conversions %)로 서비스와 리시브에 관련된 변수이다. 특히, Y7(winners including service)이 왼쪽에서, Y2(aces)와 Y11(total points won)이 오른쪽에서의 눈에 띄게 각각 선수특성변수군 중에서 (랭킹, 체중)과 신장과 각각 연관성이 있음을 보여주고 있다.

다음으로 패배한 선수들만의 자료에 대한 정준상관 행렬도 그림 3.2(b)에서 Dim1(제1축)에 대하여 Y1(1st serve %), Y8(receiving points won), Y9(break point conversions %), Y10(net approaches %), Y11(total points won)는 왼쪽에 놓여있어 주로 리시브와 관련된 변수로 해석할 수 있으며 이와 반대로 놓여진 나머지 변수는 Y4(unforced errors), Y5(winning % on 1st serve), Y6(winning % on 2nd serve), Y7(winners including service)로 주로 서비스와 관련된 득점변수이다. Y2(aces)와 Y3(double faults)은 원점 근처에 있어 어느 쪽과도 연관성을 설명하기에는 부족함을 보여준다. 특히, Y1(1st serve %)이 승패를 떠나 선수특성요인으로 랭킹과 체중에 연관성이 매우 높은 경기요인변수임을 알 수 있다.

다음으로 표 3.1은 승패별 선수특성요인과 경기요인에 대한 정준상관 행렬도의 형상 비교를 위한 프로크러스티즈 통계량을 제공하고 있다. 수치상 승패별 선수특성요인과 경기요인 정준상관 행렬도는 서로 완전히 일치 하지는 않지만 유사한 경향을 나타낸다고 여겨진다. 이는 앞서 행렬도 분석에서 지적한 내용과 일치한다.

3.3. 코트별 비교

그랜드슬램대회가 열리는 경기코트는 크게 세 종류로 주최국에 따라 미국과 호주가 하드코트, 프랑스가 앙투카코트, 영국이 잔디코트를 사용하고 있다. 이에 따른 선수특성요인과 경기요인간의 연관성 정준상관 행렬도를 통해 살펴보기로 하자. 그림 3.3은 (a) 하드코트(전미 + 전호) (b) 앙투카코트(전불) (c) 잔디코트(전영) 별 경기요인변수군에 대한 정준상관 행렬도를 제공하고 있다. 물론 선수특성요인변수군에 대한 정준상관 행렬도에서는 3.1절에서 설명한바와 같이 원편에 놓여진 체중과 랭킹이 서로 연관성이 높고, 그 반대편에는 키가 다른 경향을 보이는 변수이다.

먼저 전미와 전호 그랜드슬램대회의 하드코트에 출전한 선수들의 경기요인변수군 대한 정준상관 행렬도 그림 3.3(a)에서 Dim1(제1축)에 대하여 Y1(1st serve %), Y5(winning % on 1st serve), Y7(winners including service), Y8(receiving points won), Y11(total points won)는 왼쪽에 놓여있어 서비스를 잘 넣어 이길 확률과 관련된 변수로 해석할 수 있으며 이와 반대로 놓여진 나머지 변수는 Y2(aces), Y4(unforced errors), Y6(winning % on 2nd serve), Y9(break point conversions %), Y10(net approaches %)로 서비스와 리시브에 관련된 변수이다. 특히, Y3(double faults)는 원점 근처에 있어 어느 쪽과도 연관성을 설명하기에는 부족함을 보여준다.

다음으로 앙투카코트(붉은 벽돌가루코트)의 경기요인변수군의 정준상관 행렬도 그림 3.3(b)에서 Dim1(제1축)에 대하여 Y1(1st serve %), Y2(aces), Y6(winning % on 2nd serve), Y7(winners including service), Y9(break point conversions %), Y10(net approaches %), Y11(total points won)는 왼쪽에

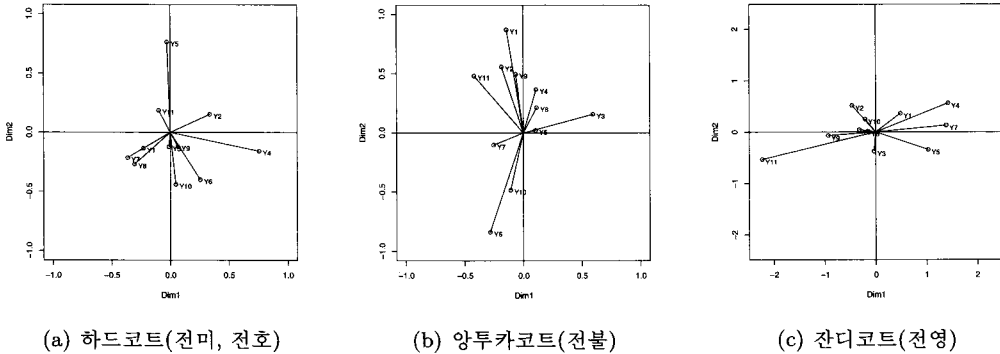


그림 3.3. 코트별 경기요인변수군에 대한 정준상관 행렬도

표 3.2. 경기코트별 행렬도상 영향력이 있는 경기요인변수 순서

하드코트	양투카코트	잔디코트
Y5 : <u>Winning % on 1st serve</u>	Y6 : <u>Winning % on 2nd serve</u>	Y11: <u>Total points won</u>
Y4 : <u>Unforced errors</u>	Y1 : <u>1st serve %</u>	Y4 : <u>Unforced errors</u>
Y10: <u>Net approaches %</u>	Y11: <u>Total points won</u>	Y7 : <u>Winners including service</u>
Y6 : <u>Winning % on 2nd serve</u>	Y3 : <u>Double faults</u>	Y5 : <u>Winning % on 1st serve</u>
Y7 : <u>Winners including service</u>	Y10: <u>Net approaches %</u>	Y9 : <u>Break point conversions %</u>
Y8 : <u>Receiving points won</u>		

놓여있어 서비스를 잘 넣어 이길 확률과 득점에 관련된 변수로 해석할 수 있으며 이와 반대로 놓여진 나머지 변수는 Y3(double faults), Y4(unforced errors), Y5(winning % on 1st serve), Y8(receiving points won)은 주로 개인 실책에 관련된 변수이다.

끝으로 잔디코트인 전영 그랜드슬램대회에 출전한 선수의 경기요인변수군 대한정준상관 행렬도 그림 3.3(c)에서 Dim1(제1축)에 대하여 Y2(aces), Y6(winning % on 2nd serve), Y9(break point conversions %), Y10(net approaches %), Y11(total points won)는 왼쪽에 놓여있어 상대의 공격에 대한 리시버와 득점에 관련된 변수로 해석할 수 있으며 이와 반대로 놓여진 나머지 변수는 Y1(1st serve %), Y4(unforced errors), Y5(winning % on 1st serve), Y7(winners including service)은 서비스를 잘 넣어 이길 확률과 주로 개인 실책에 관련된 변수이다. 특히, Y3(double faults)과 Y8(receiving points won)은 원점 근처에 있어 어느 쪽과도 연관성을 설명하기에는 부족함을 보여준다.

지금까지 경기코트별 선수특성요인변수군과 연관된 경기요인변수군에 대한 정준상관 행렬도를 살펴보았다. 끝으로 경기코트별 정준상관 행렬도에서 영향력이 있는 경기요인변수를 순서에 따라 정리한 것이 표 3.2이다. 일반적으로 행렬도에서 변수들의 좌표점은 원점으로부터 시작되는 벡터로 표시하는 데 실제로 지금까지 살펴 본 여러 행렬도에서 원점에서 시작하는 화살표에 해당한다. 이 화살표의 길이가 해당 변수의 근사적 표준편차를 나타낸다. 따라서 표 3.2에서 영향력이 있는 변수란 행렬도에서 화살표의 길이가 긴 변수를 의미한다.

표 3.2에서 먼저 경기코트별로 아래 줄 친 경기요인변수는 행렬도상 원점에 있어 선수특성요인변수 (랭킹, 체중)와 연관성이 높다. 하드코트와 잔디코트의 Y4(Unforced errors)가 두 번째 영향력 있는 변수인 것은 두 코트의 특징이 양투카 코트보다 바운드후의 볼이 빠르다는 점을 감안할 때 선수들이 공격

표 3.3. 코트별 행렬도 형상비교를 위한 프로크러스티즈 통계량

	선수특성요인			경기요인		
	하드	양투카	잔디	하드	양투카	잔디
하드	0			0		
양투카	1.322	0		4.678	0	
잔디	1.426	1.022	0	12.811	13.597	0

적인 샷을 구사하여 개인 실책이 많은 것으로 생각된다. 다음으로 하드코트와 양투카 코트의 Y10(Net approaches %)의 경우 상대와의 랠리 중에 짧은 볼이 왔을 때 넷플레이를 위한 접근이 많은 것으로 양투카코트에도 나타나지만 하드코트가 더 영향력이 있는 것으로 나타난 것은 코트의 특징상 바운드가 크고 바운드된 볼이 빠르기 때문에 코트특성에 따른 전체 선수들의 플레이스타일이 공격적인 스트로크를 구사하고 있음을 알 수 있고 이러한 코트특성이 선수들의 플레이스타일에도 영향을 미쳤다고 생각된다. 그러나 잔디코트에서는 볼이 바운드 후에 낮고 빠른 코트특징이 있어 주로 남자 선수인 경우 포인트를 결정하기 위한 서브 앤드 발리 플레이가 많으며 하드나 양투카보다 상대와의 랠리가 일반적으로 길지 않기 때문으로 생각된다. 하드코트의 Y8(Receiving point won)과 잔디코트의 Y9(Break point conversions %)는 빠른 코트에서 승리하기 위해서는 특히 리턴경기가 승패에 결정적인 역할을 하기 때문에 각각 영향력이 있는 것으로 생각된다. 끝으로 서비스와 관련된 Y3(Double faults), Y5(Winning % on 1st serve), Y6(Winning % on 2nd serve)의 경우 모든 코트에서 나타나고 있고 다른 변수들은 코트의 특성 즉 볼 바운드 후의 변화에 따른 선수들의 플레이 스타일에 따라 경기요인변수의 영향력이 나타난다고 볼 수 있다.

다음으로 표 3.3은 코트별 선수특성요인과 경기요인에 대한 정준상관 행렬도의 형상비교를 위한 프로크러스티즈 통계량이다. 먼저 선수특성요인의 경우 승패별과 같이 정준상관 행렬도가 대동소이한 결과를 제공한 것처럼 코트별 형상비교에서도 프로크러스티즈 통계량 값에서 서로 간에 유사한 결과를 제공하고 있다. 이와는 다르게 경기요인 정준상관 행렬도는 하드와 양투카 코트에 대한 4.6783보다 하드와 잔디 12.811과 잔디와 양투카 13.597 보다도 약 3배 정도 작다. 이는 상대적으로 하드와 양투카 코트의 정준상관 행렬도의 형상이 다소 비슷한 경향을 나타낸다고 여겨진다. 이는 앞서 행렬도 분석에서 잔디코트가 나머지 다른 코트와 다른 성향을 많이 보여주는 것과 관련이 있음을 의미한다.

4. 결론

세계 4대 테니스 그랜드슬램대회에 참가한 선수들의 특성요인변수군과 경기요인변수군간의 정준상관 행렬도는 이들 변수군간의 연관성을 그림으로 잘 보여 주었다. 특히, 승패별 분석에서는 승패를 떠나 특성변수군에서 랭킹과 체중에 경기요인변수군에서 Y1(1st service %)과 연관성이 높아 첫 서비스 확률이 랭킹에 중요한 변수임을 보여주었다. 경기코트별 분석에서는 하드코트와 양투카코트가 코트성격상 상대와의 랠리 중 넷 플레이로 인한 접근이 많은 반면에 잔디코트는 볼이 바운드 후에 낮고 빠른 코트특징이 있어 주로 남자 선수인 경우 포인트를 결정하기 위한 서브 앤드 발리 플레이가 많아 이와 관련된 경기요인이 많이 작용함을 알 수 있다. 그리고 이러한 차이점은 프로크러스티즈 분석을 이용한 정준상관 행렬도의 형상비교에서도 유사한 경향을 보여 주었다.

참고문헌

김도연, 최태훈 (2000). 테니스 그랜드슬램 대회의 경기분석, <한국체육측정평가학회지>, 3, 13-25.

- 최용석 (2006). <행렬도 분석, 기초과학 총서 2권>, 부산대학교 기초과학연구원.
- 최용석, 현기홍 (2006). <통계적 형상분석의 이해와 응용 -프로크러스티즈 분석의 저항성 버전 연구 및 개발->, 자유아카데미, 서울.
- 최용석, 현기홍, 정수미 (2005a). 다변량 분산분석에서 추정된 모수행렬의 행렬도, *Journal of The Korean Data Analysis Society*, **7**, 851-858.
- 최용석, 현기홍, 정수미 (2005b). MANCOVA biplot, <한국통계학논문집>, **12**, 705-712.
- 최용석, 강창완, 김경덕 (2005c). 명목형 다항반응 로지스틱회귀모형의 행렬도 분석, *Journal of the Korean Data Analysis Society*, **7**, 839-849.
- 최태훈 (2005). <테니스선수의 체격, 체력, 기초기술과 랭킹간의 관계>, 박사학위논문, 국민대학교.
- 최태훈, 최용석 (2008). 정준상관 행렬도와 군집분석을 응용한 KLPGA 선수의 기술과 경기성적요인에 대한 연관성 분석, <응용통계연구>, **21**, 429-439.
- 허명희 (1993). <統計相談의 이해>, 자유아카데미, 서울.
- Choi, Y. S. (1991). Resistant Principal Component Analysis, Biplot and Correspondence Analysis, 박사학위 논문, 고려대학교.
- 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.
- Choi, Y. S. and Kim, H. Y. (2008). Applications of cluster analysis in biplots, *The Korean Communications in Statistics*, **15**, 65-76.
- Gabriel, K. R. (1971). The biplot graphics display of matrices with applications to principal component analysis, *Biometrika*, **58**, 453-467.
- Park, M. and Huh, M. H. (1996a). Canonical correlation biplot, *The Korea Communications in Statistics*, **3**, 11-19.
- Park, M. and Huh, M. H. (1996b). Quantification plots for several sets of variables, *Journal of the Korea Statistical Society*, **25**, 599-601.

A Study on the Relationship between Player Characteristic Factors and Competitive Factors of Tennis Grand Slams Competition Using Canonical Correlation Biplot and Procrustes Analysis

Tae-Hoon Choi¹ · Yong-Seok Choi² · Sang Min Shin³

¹Department of Physical Education, Andong Science College;

²Department of Statistics, Pusan National University;

³Department of Statistics, Pusan National University

(Received April 2009; accepted May 2009)

Abstract

Canonical correlation biplot is 2-dimensional plot for investigating the relationship between two sets of variables and the relationship between observations and variables in canonical correlation analysis graphically. Recently, Choi and Choi (2008) suggested a method for investigating the relationship between skill and competition score factors of KLPGA players using canonical correlation biplot and cluster analysis. analysis. Procrustes analysis is very useful tool for comparing shape between configurations. Therefore, in this study, we will provide a method for investigating the relationship between player characteristic factors and competitive factors of tennis grand slams competition using Canonical correlation biplot and Procrustes analysis.

Keywords: Canonical correlation analysis, biplot, tennis grand slams competition, Procrustes analysis.

²Corresponding Author: Professor, Department of Statistics, Pusan National University, Busan 609-735, Korea. E-mail: yschoi@pusan.ac.kr