응용통계연구 제 8권 1호
1995년 3월 pp. 51-60.

'82-92 한국 프로야구의 각 팀과 부문별 평균 성적에 대한 추가적 주성분분석의 응용

최 용 석[1], 심 희 정[2]

요 약

크기가 $n \times p$인 자료행렬에서 p개의 변수들과 성격이 다소 다른 p_s개의 변수를 같이 고려한 크기가 $n \times (p+p_s)$ 자료행렬이 있다 하자. 전통적 주성분분석은 성격이 다른 변수들로 인하여 효과적인 결과를 제공하지 못한다. 본 논문에서는 이런 점을 개선하기 위해서 성격이 다른 p_s개의 변수를 추가변수로 두는 추가적 주성분분석을 소개하려 한다. 이 기법은 전통적 주성분분석의 대수적·기하적인 면을 따른다. 그리고 전통적 주성분분석과 추가적 주성분분석을 활용한 한국 프로야구의 8개팀과 1982-1992년 동안의 14개의 부문별 기록에 대한 전형적인 자료분석의 한 예를 제시한다. 더불어 두 분석의 결과도 비교하였다.

1. 서 론

한국 프로야구는 1982년부터 시작하여 1992년까지 지난 11년 동안 각 팀과 선수들 개인의 기록면에서 변화와 발전을 거듭해 왔다. 그러나 그동안의 기록들을 토대로 하여 이루어진 통계적 분석은 선수들 개인의 기록을 토대로 회귀분석을 이용하여 연봉계산모형을 제시하였던 윤여관(1990)을 제외하고는 거의 없다고 해도 과언이 아니다.

본 논문에서는 한국 프로야구의 8개팀(롯데, 빙그레, 삼성, 쌍방울, OB, LG, 태평양,해태)과 11년 동안(1982-1992)의 14개의 부문별 기록(도루,득점, 방어율,병살타, 사사구, 삼진, 수비율, 실책, 안타, 장타율, 출루율, 타율, 타점, 홈런)간의 관계를 설명하기 위하여 주성분분석을 이용하였다. 특히, 주성분분석에서 성격이 전혀 다른 변수들로 인하여 효과적인 분석을 제공받을 수 없다. 그러므로 성격이 다른 변수를 추가변수로 두는 추가적 주성분분석을 이용하여 이런 문제점을 해결하고자 한다.

2. 추가적 주성분분석

2.1 주성분분석

주성분분석(principal component analysis)은 차원축소를 통하여 저차원상에서 변수의 관계를 규명하는 다변량 자료분석기법이다. 최근에는 관측치와 변수를 동시에 저차원상에 나타내어

1) (609-735) 부산시 금정구 장전동 부산대학교 통계학과 조교수.
2) (609-735) 부산시 금정구 장전동 부산대학교 통계학과 석사과정 졸업.

이들의 관계를 시각적으로 설명하는 주성분분석도 있다 (Gabriel, 1971).

주성분분석을 제공하는 알고리즘에는 몇 가지가 있으나, 여기서는 비정칙치분해 (singular value decomposition)에 바탕을 둔 주성분분석을 이용하려 한다. 비정칙치분해는 다변량분석에서 차원축소를 제공하는 행렬연산 중 가장 유용한 대수기법이다 (Lebart, et al., 1984, pp. 8-9 ; Choi and Huh, 1993).

행과 열의 수가 각각 n과 p인 자료행렬을 $X=(x_{ij})$, $(i=1,...,n;\ j=1,...,p)$로 나타내고, 새로운 행렬을 $Z=\left((x_{ij}-\bar{x}_j)/s_j\right)$, $(i=1,...,n\ ;j=1,...,p)$로 정의한다. 여기서 $\bar{x}_j=\sum_{i=1}^{n}x_{ij}/n$ 이고 $s_j^2=\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2/(n-1)$이다.

그러면 자료행렬 Z의 계수(rank)가 r이라면 비정칙치분해는 다음과 같다.

$$\begin{aligned} Z &= UD_{\sqrt{\lambda}}V' \\ &= \sum_{k=1}^{r}\sqrt{\lambda_k}\,u_k\,v_k'. \end{aligned} \tag{1}$$

여기서 U는 크기가 $n\times r$이고 $U'U=I_r$ 인 직교행렬이며 V는 크기가 $p\times r$이고 $V'V=I_r$ 인 직교행렬이다. $D_{\sqrt{\lambda}}$는 크기가 $r\times r$인 대각행렬로 비정칙치 $\sqrt{\lambda_1}\geq\sqrt{\lambda_2}\geq\cdots\geq\sqrt{\lambda_r}>0$를 대각원소로 가진다.

따라서 식 (1)로부터 R^n과 R^p공간에서 차원축소된 $s(1\leq s\leq r)$ 공간상의 좌표점은 다음 두 열벡터이다.

$$A=ZV,\quad B=Z'U \tag{2}$$

일반적으로 식 (2)의 좌표점을 좌표점행렬이라 하자. 여기서 A는 n개의 관측치의 좌표점이며 B는 p개의 변수의 좌표점이다.

식 (1)의 비정칙치분해에서 제 1, 제 2 비정칙치에 대응하는 식 (2)의 두 좌표점행렬의 첫번째, 두번째 열벡터를 가지고 주성분분석그림을 그리면 이는 2차원으로 차원축소된 주성분축이 2개인 주성분분석을 시각적으로 제공한다.

일반적으로 $s(1\leq s\leq r)$차원의 주성분분석을 시도한다면 s개의 주성분축에 의해서 설명되는 분산의 정도는 전통적으로

$$\rho_s=\sum_{k=1}^{s}\lambda_k\ /\ \sum_{k=1}^{r}\lambda_k$$

이다.

2.2 추가적 주성분분석

주성분분석에서 성격이 전혀 다른 변수들로 인하여 효과적인 결과를 제공받을 수 없는 경우

가 있다. 그러므로 추가적인 기법을 필요로 하며 관측치에 대해서도 동일한 개념을 적용할 수 있다 (Lebart, et al., 1984, pp. 15-16).

크기가 $n \times p$인 자료행렬 X가 추가적인 행과 열을 가진다고 할 때, $X^+ = (x^+_{ij})$, $(i=1,...,n;\ j=1,...,p_s)$는 p_s개의 변수가 추가된 행렬이고, $X_+ = (x_{+ij})$, $(i=1,...,n_s;\ j=1,...,p)$는 n_s개의 관측치가 추가된 행렬이라고 하자. 이 추가된 행렬을 표준화하여 다음과 같은 두 행렬을 각각 얻는다.

$$Z^+ = (z^+_{ij}) = \left((x^+_{ij} - \bar{x}_j{}^+) / s^+_j \right) \tag{3}$$

$$Z_+ = (z_{+ij}) = \left((x_{+ij} - \bar{x}_j) / s_j \right) \tag{4}$$

여기서 $\bar{x}_j = \sum_{i=1}^{n} x_{ij}/n$, $s_j^2 = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2/(n-1)$, $\bar{x}_j{}^+ = \sum_{i=1}^{n} x^+_{ij}/n$,

$s_j{}^{+2} = \sum_{i=1}^{n} (x^+_{ij} - \bar{x}_j{}^+)^2/(n-1)$이다.

이 표준화된 좌표점 (3)과 (4)를 R^p와 R^n공간의 주성분축상의 α번째 투사벡터에 투영시킴으로써 이 축상의 추가좌표점을 얻고자 한다. 식 (1)로부터 R^p와 R^n공간의 주성분축상의 α 번째 투사벡터는 각각 v_α와 u_α이므로 추가적 주성분분석그림을 얻기 위한 알고리즘은 다음과 같다.

1) n_s개의 추가관측치의 좌표점과 p_s개의 추가변수의 좌표점은 각각 $Z_+ v_\alpha$, $Z^{+\prime} u_\alpha$ 이고 이를 추가좌표점벡터라 한다.

2) 1)에서 구한 $\alpha = 1, 2$ 일 때 추가좌표점벡터를 좌표점으로 하는 2차원 그림과

3) 1절에서 구한 식 (2)의 좌표점행렬의 첫번째, 두번째 열벡터를 좌표점으로 하는 2차원 그림을 겹친다.

즉, 주축이 2개인 전통적인 주성분분석의 좌표점과 추가적 기법을 이용한 주성분분석의 좌표점을 시각적으로 동시에 나타낼 수 있는 2차원의 추가적 주성분분석그림을 제공한다. 2차원의 추가적 주성분분석에서 ρ_s(분산의 정도)가 전통적 주성분분석에서 보다 항상 높아진다. 이는 단지 자료행렬의 변수의 수가 적을수록 식 (1)에서 제공되는 첫번째 고유값(비정칙치의 제곱)이 전체 고유값 합계에서 차지하는 비율은 꽤 높아지기 때문이다. 여기에 대한 자세한 설명은 Lebart, et al. (1984, 7장)을 참고 바란다. 그리고 추가적 주성분분석그림의 해석은 전통적 주성분분석그림의 해석과 동일하다.

3. 한국 프로야구의 응용

본 논문에서 사용된 자료는 한국 프로야구연감(K. B. O., 1992)을 통해 얻었다. <표 3.1>은 1982-1992년 한국 프로야구의 팀과 부문별 기록에 대한 자료이다.

현재 한국 프로야구팀은 총 8개팀(롯데, 빙그레, 삼성, 쌍방울, OB, LG, 태평양, 해태)이다. 특히 빙그레(1986년에 창단)와 쌍방울(1991년에 창단)은 11년간의 부문별 기록이 없어 <표 3.1>의 자료에서는 제외하였다. 따라서 이 자료에 대한 부문별 기록(변수)의 추가적 주성분분석을 3.1절에서 제공하려 한다. 그리고 3.2절에서 빙그레와 쌍방울, 두 팀이 <표 3.2>의 비율로 된 5개의 부문별 기록(타율, 출루율, 장타율, 방어율, 수비율)에는 포함되어 팀(관측치)의 추가적 주성분분석에서 이용되었다.

<표 3.1> 1982-1992년 한국 프로야구의 팀과 부문별 기록에 대한 자료

	타 율	출루율	장타율	안 타	홈 런	득점	타 점	도 루	사사구
롯데	0.259	0.340	0.365	10273	649	4954	4567	1113	4313
해태	0.264	0.339	0.412	10551	1078	5377	5006	1346	4200
삼성	0.273	0.348	0.405	11033	1045	5892	5464	1248	4286
OB	0.255	0.325	0.356	10162	612	4773	4427	1072	3890
태평양	0.244	0.319	0.348	9674	651	4433	4058	995	4069
LG	0.258	0.315	0.355	10353	574	4894	4496	1326	3897

	수비율	방어율	삼 진	병살타	실 책
롯데	0.977	3.69	4727	884	1072
해태	0.976	3.31	5014	1009	1102
삼성	0.979	3.74	5374	854	993
OB	0.977	3.58	4921	785	1084
태평양	0.976	4.68	5481	887	1095
LG	0.976	3.69	4738	858	1124

<표 3.2> 8개팀에 대한 비율로 표현된 5개 부문의 자료

	타 율	수비율	출루율	장타율	방어율
롯데	0.259	0.977	0.340	0.365	3.69
해태	0.264	0.976	0.339	0.412	3.31
삼성	0.273	0.979	0.348	0.405	3.74
OB	0.255	0.977	0.325	0.356	3.58
태평양	0.244	0.976	0.319	0.348	4.68
LG	0.258	0.976	0.315	0.355	3.69
빙그레	0.267	0.975	0.344	0.398	3.66
쌍방울	0.251	0.973	0.339	0.375	4.77

3.1 부문별 기록에 대한 추가적 주성분분석

한국 프로야구 6개팀(롯데, 삼성, LG, OB, 태평양, 해태)의 부문별 기록에 대해 추가적 기법을 이용한 주성분분석을 실시해 보았다. 이를 위하여 부문별 기록을 <표 3.3>과 같이 공격적, 비공격적, 수비적 성향에 따라 기본변수, 추가변수 1, 추가변수 2로 각각 나누었다.

<표 3.3> 부문별 기록의 구분

	부문별 기록	분석상 구분
변 수	타율 장타율 출루율 안타 홈런 득점 타점 도루 사사구 수비율 방어율 삼진 실책 병살타	
기본변수	타율 장타율 출루율 안타 홈런 득점 타점 도루 사사구	공격적
추가변수1	수비율 방어율	수비적
추가변수2	삼진 실책 병살타	비공격적
관측치	롯데 삼성 LG OB 태평양 해태	

기본변수는 공격적 성향인 타율, 장타율, 출루율, 안타, 홈런, 득점, 타점, 도루, 사사구이고, 추가변수 1은 수비적 성향인 방어율과 수비율로 이루어져 있다. 그리고 비공격적인 성향인 삼진, 병살타, 실책을 추가변수 2로 하였다.

실제 변수의 추가적 주성분분석을 경우 1), 경우 2)로 나누었다.

경우 1) 기본변수와 추가변수 1(수비율, 방어율)의 주성분분석

그림 3.1(a)는 전통적 주성분분석그림으로 제 1주축과 제 2주축의 고유값은 전체 고유값 합계 중에서 각각 71.2%, 14.5%로 2개의 축은 85.7%를 차지한다.

주성분분석그림에서 실선은 기본변수를, 점선은 추가변수를 나타낸다.

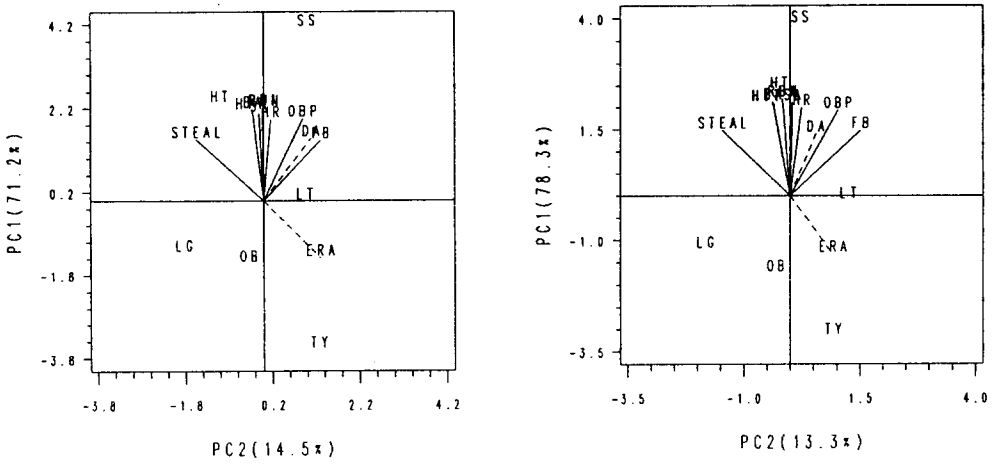
그림 3.1(a)를 살펴보면 위쪽으로 향하고 있는 BA(타율), HIT(안타), RUN(득점), RBI(타점), SA(장타율), HR(홈런)의 실선이 이루는 각이 작다. 그러므로 이들은 상관관계가 높아 비슷한 성격의 부문별 기록이라는 것을 의미한다. 특히 그 중에서도 HIT(안타)와 BA(타율), RUN(득점)과 RBI(타점)이 상관관계가 더 높음을 알 수 있다.

제 1주축(세로축)에 대해 윗쪽에 위치한 강한 공격력의 HT(해태), SS(삼성)이 BA(타율), HIT(안타), RUN(득점), RBI(타점), OBP(출루율), HR(홈런), SA(장타율)쪽으로 향하고 있다. 반면 반대편인 LG, OB, TY(태평양)은 공격력이 약하다는 것을 알 수 있다. 그리고 오른편의 윗쪽에 위치한 FB(사사구)는 축의 윗쪽인 LT(롯데), SS(삼성), HT(해태)가 높게 나타난다.

제 2주축(가로축)에 대해서는 DA(수비율)은 축의 오른쪽에 위치한 SS(삼성), TY(태평양), LT(롯데)로 향해 있다. 그리고 ERA(방어율)의 방향으로 향해 있는 TY(태평양), LT(롯데), SS(삼성)의 높은 방어율 집단과 반대로 안정된 HT(해태), LG, OB로 나눌 수 있다.

추가적 주성분분석그림 3.1(b)는 3.1(a)와 비교해 볼 때 큰 변화가 없다. 그러나 점선으로 표현된 DA(수비율)과 ERA(방어율)이 이루는 각이 더 커졌으며 실제 상관계수가 -0.153으로써 약간의 음의 상관관계를 나타낸다.

제 1주축과 제 2주축은 고유값 합계 중 각각 78.3%, 13.3%를 차지해 전체의 약 91.6%를 설명한다.



(a) 주성분분석그림 (b) 추가적 주성분분석그림

* 팀 : LT(롯데), SS(삼성), LG, OB, TY(태평양), HT(해태)
* 부문 : BA(타율), SA(장타율), DA(수비율), ERA(방어율), OBP(출루율), HIT(안타), HR(홈런), RUN (득점), RBI(타점), STEAL(도루), FB(포볼)

그림 3.1 기본변수와 추가변수 1 (수비율, 방어율)을 고려한 주성분분석

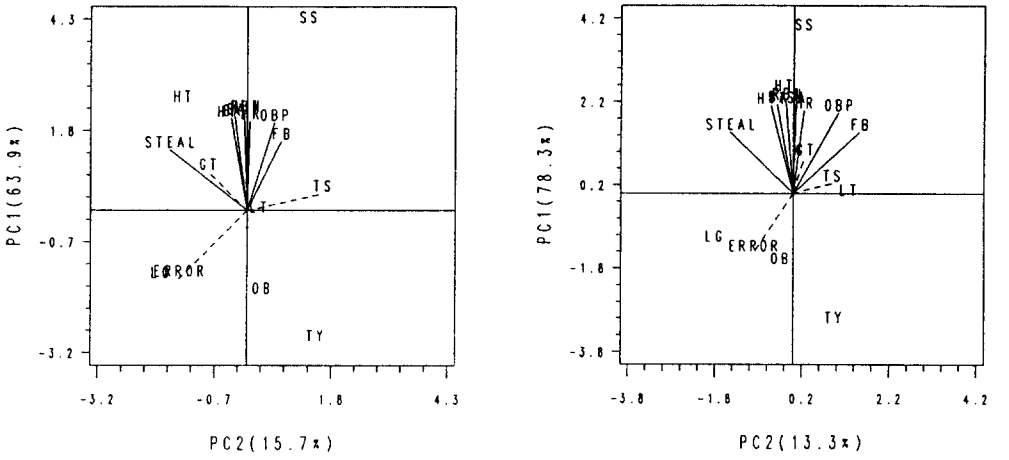
경우 2) 기본변수와 추가변수 2(삼진, 실책, 병살타)의 주성분분석

그림 3.2(a)는 기본변수와 추가변수 2를 모두 합한 경우이다. 제 1, 제 2주축의 고유값은 고유값의 합계 중에서 각각 63.9%, 15.7%를 차지하여 2차원의 주성분분석이 원 자료를 약 79.5% 정도 근사시키고 있음을 알 수 있다.

공격적인 부문에 대한 주성분분석그림의 해석은 그림 3.1(a)와 대동소이하다. 제 2주축(가로축)에 대해선 축의 왼쪽에 위치한 HT(해태)와 LG가 STEAL(도루)와 ERROR(실책)에 대해 높게 나타났으며 오른쪽에 위치한 SS(삼성), TY(태평양), OB, LT(롯데)는 낮은 경향을 나타낸다. 축의 오른쪽의 TS(삼진)에 대해서는 SS(삼성), TY(태평양), OB를 하나의 집단으로 구분할 수 있다. 삼진은 선구안이 나쁜 타자들이나 장타가 많은 타자들이 스윙폼이 커짐에 따라 많이 당하는 수가 있다. 이런 이유에서 삼성, 해태, 태평양이 삼진이 많은 경향이 있다.

추가적 주성분분석그림 3.2(b)에서 HR(홈런)과 SA(장타율)이 이루는 각이 주성분분석그림에서 보다 훨씬 작아져 두 변수간의 높은 상관관계를 잘 설명한다고 볼 수 있다. 그리고 점선으

로 나타난 GT(병살타), TS(삼진)이 이루는 각도 작아졌다. 이것은 비공격적 부문인 병살타와 삼진은 약간은 양의 관련성이 있는 기록임을 보여준다. 그리고 GT(병살타)와 공격적 부문(SA(장타율), HR(홈런))의 상관관계도 높아졌다. 이는 공격력이 높은 팀이 병살타도 많이 기록하는 경향이 있음을 나타낸다.



(a) 주성분분석그림 (b) 추가적 주성분분석그림

* 팀 : LT(롯데), SS(삼성), LG, OB, TY(태평양), HT(해태)
* 부문 : BA(타율), SA(장타율), TS(삼진), ERROR(실책), GT(병살타), OBP(출루율), HIT(안타), HR(홈런), RUN(득점), RBI(타점), STEAL(도루), FB(포볼)

그림 3.2 기본변수와 추가변수 2(삼진, 병살타, 실책)를 고려한 주성분분석

3.2 팀에 대한 추가적 주성분분석

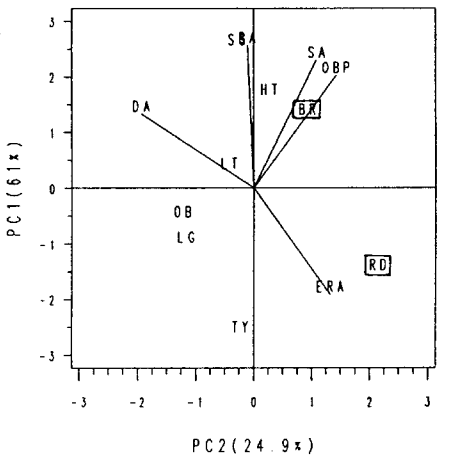
비율로 표현된 5개의 부문별 기록(타율, 방어율, 수비율, 장타율, 출루율)에 대해 2팀(빙그레, 쌍방울)을 추가관측치로 하여 팀에 대한 추가적 주성분분석을 실시하였다. 변수와 관측치에 대한 구분은 <표 3.4>에 정리해 두었다.

<표 3.4> 관측치의 구분

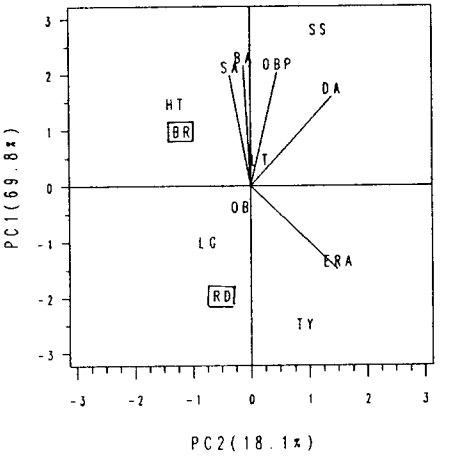
변 수	타율 방어율 수비율 장타율 출루율
관측치	롯데 삼성 LG OB 태평양 해태
추가관측치	빙그레 쌍방울

그림 3.3(a)는 8개팀 전체에 대한 분석이며 그림 3.3(b)는 빙그레와 쌍방울을 추가관측치로 둔 것이다. 먼저 그림 3.3(a)에서 제 1주축과 제 2주축은 전체 고유값의 합계 중 각각 61%, 24.9%를 차지하여 전체의 85.9%를 차지한다.

그림에서 사각형으로 표시된 것이 추가관측치이다. 제 1주축에 대해서 BA(타율), SA(장타율), OBP(출루율)에 대해 그림의 윗쪽인 SS(삼성), HT(해태), LT(롯데), BR(빙그레)와 아랫쪽의 OB, LG, TY(태평양), RD(쌍방울)로 공격적인 면에서 구분이 될 수 있다. 제 2주축에 대해선 DA(수비율)의 실선이 향해 있는 그림 3.3(a)의 왼쪽과 그 반대쪽인 오른쪽으로 구분되어 HT(해태), BR(빙그레), RD(쌍방울)은 수비에 있어 불안하며, SS(삼성), LT(롯데), OB, LG는 수비가 안정적이라는 것을 알 수 있다.



(a) 주성분분석그림



(b) 추가적 주성분분석그림

* 팀 : LT(롯데), BR(빙그레), SS(삼성), RD(쌍방울), LG, OB, TY(태평양), HT(해태)
* 부문 : BA(타율), SA(장타율), DA(수비율), ERA(방어율), OBP(출루율)

그림 3.3 팀에 대한 추가적 주성분분석

다음 추가적 주성분분석그림인 3.3(b)는 SA(장타율), BA(타율), OBP(출루율)의 실선이 이루는 각이 3.3(a)에서 보다 작아졌다. 특히 BA(타율)과 OBP(출루율), BA(타율)과 SA(장타율)의 실선이 이루는 각이 작아졌다. 타율과 출루율은 둘 다 안타의 수에 많은 영향을 받으므로 높은 상관관계가 있다고 볼 수 있다. 반면에 윗쪽의 오른편으로 실선이 향해 있는 DA(수비율)과 ERA(방어율)은 이루는 각이 직각에 가까우므로 관련이 없음을 보여준다. 제 2주축에 대해서 그림 3.3(a)와 3.3(b)를 비교해 볼 때 BR(빙그레), RD(쌍방울), HT(해태)가 그림 3.3(a)에서는 ERA(방어율)쪽으로 향하였다. 그런데 3.3(b)에서는 BR(빙그레), HT(해태)는 ERA(방어율)의 반대쪽에 위치하고 있으며 LT(롯데), SS(삼성), TY(태평양)은 ERA(방어율)쪽으로 향하고 있어 높은 방어율을 나타내고 있다.

제 1주축과 제 2주축이 전체 고유값 합계 중 각각 69.8%, 18.1%를 차지해 원자료에 87.9% 정도로 근사되고 있다.

4. 결 론

본 소고에서는 한국 프로야구의 8개 팀과 14개 부문의 기록에 대한 통계적 자료분석을 위하여 주성분분석과 추가적 주성분분석을 활용하였다.

추가적 주성분분석이 전통적 주성분분석 보다 변수들의 상관관계를 높여줌으로써 비슷한 성격의 변수들을 잘 설명해 주었다.

경우 1)과 경우 2)의 추가적 주성분분석에서 나타난 뚜렷한 결과를 몇가지 요약해 보자.

경우 1)에서 살펴보면 주성분분석 보다 추가적 주성분분석에서 방어율과 수비율이 더 큰 각을 이룬다. 이것은 이들 기록들이 음의 상관을 가진 반대 성격임을 보여준다. 경우 2)에서는 추가적 주성분분석그림에서 병살타의 위치가 많이 변하였다. 병살타와 삼진이 전통적 주성분분석에서 보다 큰 양의 상관관계를 나타낸다. 이는 비공격적인 부문에 관한 두 기록의 유사한 성격을 잘 설명해 주고 있다. 또한 전통적 주성분분석보다 병살타와 공격적 부문(홈런,장타율)이 이루는 각이 작아져 공격력이 높은 팀이 병살타도 많이 기록하는 경향을 나타낸다. 그리고 장타율과 홈런이 이루는 각이 주성분분석그림보다 더 작아져 높은 상관관계를 설명해 준다.

덧붙여 부문별 기록에 대한 각 팀의 연도별 추세분석도 고려해 볼 수 있겠다. 이에 대해서는 추후의 발표에서 이루어질 것이다.

앞으로 프로야구 기록에 대한 다양한 통계적 분석이 있으리라 생각된다. 다만 본 소고에서 프로야구의 각 팀과 부문별 기록에 대한 통계적 분석의 한 전형을 제시함에 의의를 두고자 한다.

5. 감사의 글

본 논문을 심사해 주시고 조언해 주신 심사위원께 깊은 감사를 드립니다. 그리고 2차원 그림에서 실선을 그을 수 있는 방법을 가르쳐 주신 고려대학교 허명회 교수님께 고마움을 표합니다.

참고문헌

[1] 윤여관 (1990). 『한국 프로야구 연봉자료에 관한 통계적 분석』, 석사학위논문, 고려대학교.

[2] K. B. O. (1992). 『한국 프로야구연감』.

[3] Choi, Y. S. and Huh, M. H. (1993). Resistant Singular Value Decomposition and its Applications, *Unpublished.*

[4] Gabriel, K. R. (1971). The biplot graphics display of matrices with applications to principal component analysis, *Biometrika,* Vol. 58, 453-467.

[5] Lebart, L., Morineau A. and Warwick, K. M. (1984). *Multivariate Descriptive Statistical Analysis : Correspondence Analysis and Related Techniques for Large Matrices,* Wiley, New York.

Applications of the Supplementary Principal Component Analysis for the 1982-1992 Korean Pro Baseball Data

Yong-Seok Choi[1] and Hee-Jeong Shim[1]

Abstract

Given an $n \times p$ data matrix, if we add the p_s variables somewhat different nature than the p variables to this matrix, we have a new $n \times (p+p_s)$ data matrix. Because of these p_s variables, the traditional principal component analysis can't provide its efficient results. In this study, to improve this problem we review the supplementary principal component analysis putting p_s variables to supplementary variable. This technique is based on the algebraic and geometric aspects of the traditional principal component analysis. So we provide a type of statistical data analysis for the records of eight teams and fourteen fields of the 1982-1992 Korean Pro Baseball Data based on the supplementary principal component analysis and the traditional principal component analysis. And we compare the their results.

1) Dept. of Statistics, Pusan National University, Pusan, 609-735, KOREA.