

한국 프로야구 타자의 경기력요인 분석

Analysis of Professional Korean Baseball Batters Performances Factors

저자 (Authors)	양도업, 조은형, 배상우, 정상원 Yang, Do-Eob, Cho, Eun-Hyung, Bae, Sang-Woo, Jung, Sang-Won
출처 (Source)	한국사회체육학회지 60 , 2015.5, 305-313(9 pages) Journal of Sport and Leisure Studies 60 , 2015.5, 305-313(9 pages)
발행처 (Publisher)	한국사회체육학회 KOREAN SOCIETY OF SPORT AND LEISURE STUDIES
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06395079
APA Style	양도업, 조은형, 배상우, 정상원 (2015). 한국 프로야구 타자의 경기력요인 분석. 한국사회체육학회지, 60, 305-313
이용정보 (Accessed)	이화여자대학교 203.255.***.68 2020/08/29 10:54 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

한국 프로야구 타자의 경기력요인 분석

양도업 (고려대학교) · 조은형 (성균관대학교) · *배상우 (고려대학교) · 정상원 (진주교육대학교)

I. 서 론

1. 연구의 필요성 및 목적

현대 야구는 미국, 일본, 한국, 중앙아메리카를 비롯한 많은 국가에서 인기 있는 종목으로 그동안 누적된 수많은 기록을 통해 통계적으로 신뢰할 수 있는 다양한 분석을 가능케 하고 있다(이장택, 2014a; Philip, B., Paul D. B., Bruce D. W., 2012). 특히, 기존의 단순한 통계량을 가공하여 야구기록을 보다 수학적 · 과학적으로 분석하는 세이버메트릭스(sabermetrics) 분야의 중요성은 점차 강조되고 있으며(강정구, 박석천, 김종현, 2014; 조영석, 조용주, 신상근, 2007), 이는 타율, 타점, 득점 등과 같이 전통적인 방법으로 선수의 기량을 분석하는 것에서 벗어나 더 객관적이고 구체적이며 다양한 방법으로 선수의 경기력을 평가할 수 있다는 점에서 야구경기 분석의 주류로 자리 잡아 가고 있다(Bill James, 2001; 송희배, 강기훈, 2012).

여기서 세이버메트릭스는 미국야구연구협회(Society of American Baseball Research)와 메트릭스(metrics)의 합성어로 1980년대 초 빌 제임스(Bill James)에 의해 창

안되었으며, 초기에는 야구계에 별다른 관심을 불러일으키지 못하였으나 2000년대 초 오클랜드 애슬레틱스 빌리 빈 단장의 세이버메트릭스를 기반으로 한 데이터 야구가 4년 연속 포스트시즌 진출과 20연승이라는 대기록을 달성함에 따라 주목받기 시작하였다. 최근에는 팬그래프(www.fangraphs.com), 베이스볼레퍼런스(www.baseball-reference.com)와 같은 전문적인 야구 통계 사이트가 활발히 운영되면서 많은 학자들이 다양한 세이버메트릭스 지수를 개발하여 이를 공유하고 있다. 또한 세이버메트릭스 지수는 연구논문뿐만 아니라 방송, 신문과 같은 대중매체에서도 쉽게 접할 수 있으며, 야구라는 종목에 국한되지 않고 농구, 축구, 테니스 등으로 활용범위를 점차 넓혀가고 있기 때문에 이를 제대로 알고 이해하는 것은 현대 스포츠를 즐기는데 있어 중요한 지표가 될 수 있다(원경주, 2008).

야구는 종목의 특성상 경기에서 이루어지는 선수의 거의 모든 행위가 부호화되기 때문에 통계적으로 분석이 용이하여 그동안 많은 학자들에 의해 분석되었는데, 최근에는 세이버메트릭스 지수를 이용한 연구도 활발히 이루어지고 있으며 이러한 연구들을 살펴보면 이장영(2001)과 박승현(2008)은 한국 프로야구 타

본 연구는 고려대학교 특별연구비에 의하여 수행되었음.

* sbae@korea.ac.kr

자의 연봉에 영향을 미치는 경기력을 분석한 결과 출루율, 선수경력, 타점, 경기수 등이 밀접한 영향을 미치고 있음을 나타냈고, 승희배, 강기훈(2012)은 RC(Runs Created), XR(Extrapolated Runs), BR(Batting Runs), SECA(Secondary Average) 등과 같은 세이버메트릭스 지수를 이용하여 구단별 타자의 경기력과 연봉 간의 관계를 분석하였다. 또한 이장택(2014b)의 경우 기존의 국내연구에서 사용하지 않았던 GPA(Gross Production Average), wOBA(Weighted On Base Average)와 같은 세이버메트릭스 지수를 이용하여 타자의 능력을 측정하였다. 또한 Philip B.와 2인(2012)은 타자의 득점생산과 투수의 실점예방을 예측하기 위하여 타율, 도루, 홈런수와 같이 전통적으로 사용되어온 기록과 OPS+, wOBA, ISO(Isolated Power), UZR(Ultimate Zone Rating)등의 세이버메트릭스 지수 기록으로 구분하여 비교분석하였다.

대부분의 국내 연구는 KBO에서 제공되는 자료의 한계로 인하여 다양한 세이버메트릭스 지수를 활용 할 수 없는 한계점을 지니고 있으며, 시대별로 다른 기록의 가치를 동일하게 취급하여 발생하는 오류와 세이버메트릭스 지수를 계산할 때 사용되는 상수 값이 미국 프로야구 데이터를 기반으로 사용되고 있다는 것을 공통된 단점으로 지적하며 이를 개선하기 위한 후속연구가 반드시 필요하다고 많은 선행연구에서 제안하고 있다(이영훈, 2007; 이장택, 2014a, 2014b; 장진희, 문춘걸, 2014; 조영석, 조용주, 2005).

따라서 세이버메트릭스 지수를 이용하여 과학적이고 객관적인 방법으로 타자의 경기력요인을 분석하는 것은 선수 개인에게는 선수간의 임금불균형을 해소하고 연봉 협상 시 발생하는 구단과의 마찰을 줄일 수 있으며, 구단측면에서는 정확한 데이터를 바탕으로 다양한 작전과 전략을 통해 팀의 승리에 기여할 수 있기 때문에 반드시 필요한 것으로 사료된다. 또한 야구에 대한 대중의 관심 증대로 인해 발생하는 여러 가지 의문을 해소함으로써 한국 프로야구의 지속적인 흥행을 유지하는데 이바지할 수 있을 것으로 기대된다.

II. 연구방법

1. 연구대상

한국 프로야구 타자들의 경기력요인을 분석하기 위하여 2008~2014년 한국 프로야구 페넌트레이스 규정타석을 채운 타자 307명의 공식기록을 연구대상으로 하였다. 자료를 2008~2014년으로 선정한 것은 본 연구의 목적과 더불어 최근 한국 프로야구에서 두드러지게 나타나고 있는 타고투저의 현상을 분석하기 위한 일환으로 2008년 이후부터 통산 타율보다 더 높은 수치를 지속적으로 기록하고 있기 때문이다. 다만 2012년의 경우 통산 타율보다 낮은 수치를 기록하였지만 자료의 연속성을 고려하여 분석에 포함하였다.

자료의 객관성을 확보하기 위해 KBO(Korea Baseball Organization, 한국야구위원회)에서 매년 제공하는 한국 프로야구연감 자료를 사용하였으며, 각 팀별로 조사된 선수들의 수는 신생팀인 NC를 제외하면 롯데가 43명으로 가장 많았고 SK가 32명으로 가장 적었다.

2. 측정도구 및 방법

본 연구에서는 한국 프로야구 타자의 경기력요인을 분석하기 위해 세이버메트릭스 지수를 활용하였다. 현재 개발되어 있는 세이버메트릭스 지수는 상당수 있지만 그 중 타자의 경기력을 분석하는데 필요한 지수로 제한하였고, 체육전공 교수 1인, 야구전문가 2인의 도움을 받아 본 연구의 목적과 부합하는 세이버메트릭스 지수를 선정하였다. 자세한 내용은 다음과 같다.

① OPS(On base percentage Plus Slugging percentage, 공격공헌도)

세이버메트릭스 지수중에서 가장 보편화 된 것으로 출루율과 장타율의 합으로 계산되는 간편한 공식 때문에 타자를 평가하는데 있어 간단하면서도 매우 효과적

인 지수라고 할 수 있다. 장타율에 비해 출루율이 저평가되는 단점을 지니고 있음에도 불구하고 타자의 공격능력을 종합적으로 평가해준다.

② GPA(Gross Production Average, 총생산평균)

OPS보다 더 정교한 개념으로 장타율이 과대평가된다는 단점을 보완하여 출루율에 1.8의 가중치를 부여해 만들어진 지수이다.

③ SECA(Secondary Average, 수정타율)

안타를 타수로 나누는 타율의 공식이 갖는 가장 큰 단점은 장타와 단타의 가치를 동일 시 하면서 볼넷은 인정하지 않으므로 SECA는 장타율의 가중치에 4사구와 도루의 가치를 고려해 만든 지수이다.

④ TA(Total Average, 종합공격력)

총 진루수를 총 아웃수로 나눈 것으로 한 시즌동안 타자가 한 번의 공격기회에서 얼마만큼 진루했는가를 나

타내는 지수이다.

⑤ RC(Run Created, 득점생산력)

세이버메트릭스의 창시자인 빌 제임스가 고안한 것으로 타자의 출루능력과 주자를 진루시키는 능력을 타자의 득점에 관한 창출능력으로 보고 한 타자가 팀이 득점을 올리는데 있어서 어느 정도 기여했는지를 나타낸다.

⑥ RC/27(Run Created per 27 outs, 경기당 득점기여도)

RC가 타자가 몇 점을 만들어냈는지를 의미하는 것이라면 RC/27은 한 타자로 9명이 구성된 팀이 한 경기를 치르면 몇 점이나 뽑을 것인가를 평균수치화 한 것이다.

⑦ XR(Extrapolated Runs, 추정득점력)

RC와 유사한 개념으로 팀의 득점에 얼마나 공헌했는지를 나타내며 RC와의 차이점으로는 1955년부터 1997년까지 메이저리그 공식기록을 회귀분석하여 도출한 선형공식이라는 점으로 정확도 면에서는 RC보다 낮다고

표 1. 세이버메트릭스 지수 공식

지수	공식
OPS	OPS=출루율+장타율 출루율=(안타+4사구)/(타수+4사구+희비) 장타율={1루타+(2루타×2)+(3루타×3)+(홈런×4)}/타수
GPA	GPA={출루율×1.8+장타율}/4
SECA	SECA={2루타+(3루타×2)+(홈런×3)+4사구+도루-도실}/타수
TA	TA=(루타수+몸에 맞는 볼+볼넷+도루)/(타수-안타+도실+병살타)
RC	RC=A×B/C A=(안타+4사구-도실-병살타) B={루타수+0.52×(도루+희생타)+0.26×(4사구-고의4구)} C=(타수+4사구+희생타)
RC/27	RC/27=(RC×27)/A A=(타수-안타+희타+희비+도실+병살타)
XR	XR=(1루타×0.5)+(2루타×0.72)+(3루타×1.04)+(홈런×1.44)+{(4사구-고의4구)×0.34}+(고의4구×0.25)+(도루×0.18)-(도실×0.32)-{(타수-안타-삼진)×0.09}-(삼진×0.098)-(병살타×0.37)+(희비×0.37)+(희타×0.04)
BR	BR=(1루타×0.47)+(2루타×0.78)+(3루타×1.09)+(홈런×1.40)+{(볼넷+몸에 맞는 볼)×0.33}+(도루×0.30)-(도실×0.60)-{(타수-안타)×0.25}-(주루사×0.50)
ISO	ISO=장타율-타율
PSN	PSN=(홈런×도루×2)/(홈런+도루)
wOBA	wOBA=[{(볼넷-고의4구)×0.7}+(몸에 맞는 볼×0.73)+(1루타×0.89)+(2루타×1.27)+(3루타×1.61)+(홈런×2.07)+(도루×0.25)+(도실×0.5)]/(타석-고의4구)
OW%	OW% $=\frac{(RC/27)^2}{(RC/27)^2 + 리그(RC/27)^2}$
BABIP	BABIP=안타-홈런/타수-삼진-홈런+희비

알려져 있다.

⑧ BR(Batting Runs, 득점공헌도)

리그 타자들의 평균공격력을 0으로 놓은 상태에서 상대적으로 타자의 팀 공격기여도를 수치화 한 지수이다.

⑨ ISO(Isolated Power, 순수장타율)

타자의 파워히팅 능력을 나타낸 것으로 장타율에 타율이 포함되어 있는 것을 고려해서 고안된 지수이다.

⑩ PSN(Power Speed Number, 호타준족)

타자의 호타준족 정도를 나타낸 지수로 도루실패를 감안하지 않는다는 단점을 지니고 있다.

⑪ wOBA(Weight On Base Average, 타석 당 득점기대)

타자가 타석에 들어섰을 때의 여러 가지 상황에 따른 득점 가치를 고려하여 타자의 생산력을 나타낸 것으로 득점과의 상관관계가 매우 높고, 출루율의 가치가 저평가 받는 문제점을 개선하였지만 구장효과를 반영하지 못하며 주루와 타격을 분리하지 못했다는 단점을 지닌다.

⑫ OW%(Offensive Winning Percentage, 공격기대승률)

리그의 평균 득점과 한 타자의 RC/27을 고려해서 한 타자만으로 이루어진 타선이라면 몇 %의 승률을 가지는지를 나타내는 지수로 리그평균의 변화를 고려하여 한 타자의 한 시즌의 리그 지배력을 알 수 있다.

⑬ BABIP(Batting Average on Balls In Play, 인플레이 타구비율)

타자가 친 공이 페어지역 안에 떨어진 경우만을 나타내는 지수로 타자와 투수에게 모두 적용이 가능하다. 본인의 타격 스타일에 따라 자신만의 고유한 BABIP를 가지게 되며, 라인드라이브> 그라운드볼>플라이볼 순으로 BABIP 값이 높게 형성된다.

3. 자료처리

KBO에서 제공하는 한국프로야구연감을 토대로 연구대상자의 기록을 코딩하였고, 세이버메트릭스 지수의

수치를 입력 후 SPSS ver. 22.0을 이용하여 분석하였다. 연구 자료의 기본적인 내용을 알아보기 위해 기술통계를 실시하였고, 지수들 간의 상관관계를 알아보기 위하여 상관분석을 이용하였다. 또한 타자들의 경기력요인으로 타자의 등급을 나누기 위해서 산출된 세이버메트릭스 지수를 바탕으로 군집분석을 실시하였다. 끝으로 주성분회귀분석으로 추출된 요인점수에 세이버메트릭스 지수를 대입하여 분석하였다.

Ⅲ. 연구결과

1. 데이터분석

2008~2014년 한국 프로야구 페넌트레이스 규정타석을 채운 타자 307명의 경기력 평가를 위해 기본적인 내용을 알아보기 위해 기술통계를 실시한 결과는 <표 2>와 같다.

<표 2>의 세이버메트릭스 지수 중 변화량이 큰 측도는 RC, XR, BR으로 나타났는데 이것은 이 지수들의 계산 공식에서 보면 다른 지수들에 비해 많은 타격지표를 사용함에 따라 상대적으로 편차가 크게 나타나는 지수들

표 2. 세이버메트릭스 지수 기술통계

	평균	표준편차	최소값	최대값
OPS	.815	.105	.593	1.198
GPA	.278	.033	.210	.391
SECA	.309	.084	.155	.636
TA	.822	.150	.530	1.404
RC	74.772	21.685	37.375	146.742
RC/27	6.328	1.854	3.309	14.050
XR	88.464	23.622	45.814	162.664
BR	39.741	22.017	-2.620	114.870
ISO	.151	.063	.017	.383
PSN	7.537	5.568	.000	27.679
wOBA	.433	.057	.317	.646
OW%	.583	.127	.299	.885
BABIP	.326	.031	.214	.405

표 3. 세이버메트릭스 지수 상관행렬

	OPS	GPA	SECA	TA	RC	RC/27	XR	BR	ISO	PSN	wOBA	OW%	BABIP
OPS	1	.993**	.824**	.959**	.908**	.971**	.906**	.968**	.844**	.252**	.980**	.964**	.528**
GPA		1	.812**	.967**	.909**	.979**	.891**	.958**	.782**	.231**	.964**	.976**	.573**
SECA			1	.877**	.771**	.806**	.798**	.838**	.856**	.302**	.830**	.787**	.162**
TA				1	.931**	.979**	.909**	.960**	.772**	.321**	.954**	.958**	.536**
RC					1	.932**	.984**	.968**	.724**	.357**	.903**	.904**	.522**
RC/27						1	.899**	.955**	.752**	.237**	.944**	.965**	.588**
XR							1	.974**	.798**	.389**	.919**	.886**	.438**
BR								1	.833**	.350**	.972**	.943**	.497**
ISO									1	.275**	.860**	.730**	.112*
PSN										1	.326**	.263**	.130*
wOBA											1	.942**	.512**
OW%												1	.608**
BABIP													1

**p<.01, *p<.05

이기 때문에 분석된다. 또한 <표 3>은 13개 세이버메트릭스 지수들의 상관행렬로 BABIP와 ISO, PSN의 관계를 제외한 나머지 모든 지수들의 상관관계가 유의수준 1%에서 유의함을 나타내고 있다.

2. 군집분석

산출된 세이버메트릭스 지수를 바탕으로 타자들의 공통되는 성향을 분석하기 위해 K-평균 군집분석을 실시하였다. 군집의 수를 3개로 하였을 시에는 각 군집의 특성이 잘 나타나지 않았고, 5개로 하였을 시에는 군집의 대상이 너무 분산되고 특정 집단에 편중되는 현상이 나타나 4개의 군집으로 분류하는 것이 가장 적절한 것으로 분석되었다. 각 군집의 특성으로는 G1은 가장 우수한 집단, G2는 우수한 집단, G3는 평균수준의 집단, G4는 부족한 집단으로 분류하였다. 각 군집에 해당하는 선수의 수와 군집별 세이버메트릭스 지수의 평균비교 값은 <표 4>와 같다.

분류된 집단 간 차이의 유의성을 검증하기 위해 일원 배치분산분석 후 Duncan으로 사후분석을 실시한 결과, PSN과 BABIP에서만 G1>G2, G3>G4로 나타났고, 나머지 지수는 모두 G1>G2>G3>G4로 그룹간의 차이가 유의한 것으로 드러났다.

표 4. 군집별 세이버메트릭스 지수 평균비교

	G1(29)	G2(81)	G3(117)	G4(80)
OPS	1.017	0.888	0.793	0.700
GPA	.340	.300	.272	.243
SECA	.450	.360	.292	.229
TA	1.129	.914	.794	.660
RC	121.878	88.851	69.402	51.295
RC/27	10.250	7.399	5.911	4.433
XR	136.747	105.876	82.737	61.706
BR	85.116	55.553	34.523	14.915
ISO	.250	.195	.134	.095
PSN	11.554	9.132	7.359	4.728
wOBA	.540	.474	.422	.369
OW%	.798	.675	.572	.430
BABIP	.357	.332	.328	.304

3. 주성분회귀분석

요인분석의 두 가지 방법 중 주성분분석은 서로 관련이 있는 n개 측정변인들의 선형조합(linear combination)을 통해서 서로 독립적인 합성변인(composite variable)을 추출하는 통계기법으로 합성변인은 측정변인보다 적은 수로 n개 측정변인의 총분산을 충분하게 설명하도록 추출되기 때문에 자료의 축소나 요약에 위한 기법이다(엄한주, 2001). 이러한 주성분분석은 본 연구의 목적에

가장 부합하는 통계적 기법으로 주성분분석에 앞서 변인들의 상관행렬은 <표 3>에서 검정되었다.

<표 5>는 세이버메트릭스 지수를 요인분석한 결과로 총 13개의 지수 중 PSN과 BABIP가 낮은 요인적재값을 보여 2개의 요인을 제외한 11개 지수를 추출하였다. 고유치가 1.0 이상인 1개의 요인으로 추출되었으며, 전체 분산 설명력은 90.88%로 나타나 1개의 주성분이면 11개 세이버메트릭스 지수의 설명이 가능하다고 간주하였다.

표 5. 세이버메트릭스 지수 요인분석

지수	성분
OPS	.985
GPA	.978
TA	.981
RC	.950
RC/27	.973
XR	.951
BR	.990
wOBA	.980
OW%	.961
SECA	.875
ISO	.850

KMO=.780, Bartlett구형성검정=11774.163, df=.55, sig=.000

요인분석 후 각 요인의 요인점수를 이용하여 2차 분석, 즉 회귀분석을 실시하여 산출된 11개 세이버메트릭스 지수에 대한 표준화 계수 값은 <표 6>과 같다. 11개 세이버메트릭스 지수의 표준화된 변수가 Z1에서 Z11라고 하면 한 선수가 가지는 요인점수 S는 아래와 같은 식으로 나타낼 수 있다.

$$S = 0.099Z_1 + 0.098Z_2 + 0.087Z_3 + 0.098Z_4 + 0.095Z_5 + 0.097Z_6 + 0.095Z_7 + 0.099Z_8 + 0.085Z_9 + 0.098Z_{10} + 0.096Z_{11}$$

표 6. 요인점수의 표준화 계수

OPS	GPA	SECA	TA	RC	RC/27	XR	BR	ISO	wOBA	OW%
0.099	0.098	0.087	0.098	0.095	0.097	0.095	0.099	0.085	0.098	0.096

4. 타자의 경기력분석

주성분회귀분석을 통해 구해진 요인점수 S에 모든 선수들의 표준화된 세이버메트릭스 지수를 대입하여 구한 2008년~2014년 상위 10명은 <표 7>과 같다. 순위를 살펴보면 박병호와 최형우가 각각 두 번씩 순위에 올라있고, 2014년도에 해당하는 선수가 4명씩이나 순위에 속해있다.

2008년~2014년 중에서 가장 타고투저현상이 가장 심했던 년도를 살펴보면 2014년으로 정규시즌 9개 팀의 평균자책점(ERA) 평균이 5.21로 1982년부터 2014년까지 통산 팀 평균자책점 평균(4.07)을 훨씬 웃도는 수치이다. 또한 2014년 전체 팀 타율 평균은 0.289로 통산 팀 타율 평균(0.263)을 상회하고 있다. 소속팀은 넥센 소속의 선수가 3명으로 가장 많다.

요인점수 S로 구해진 상위 10% 타자들의 소속군집을 살펴보면 1명을 제외한 모든 선수들이 G1에 속해 있어 군집의 분류가 적절히 이루어져있음을 알 수 있다.

표 7. 요인점수 순위

순위	이름	해당연도	소속팀	요인점수
1	강정호	2014	넥센	3.615
2	박병호	2014	넥센	3.130
3	테임즈	2014	NC	2.941
4	이대호	2010	롯데	2.708
5	최형우	2011	삼성	2.454
6	박병호	2013	넥센	2.385
7	최형우	2014	삼성	2.328
8	페타지니	2009	LG	2.249
9	김현수	2009	두산	2.233
10	최희섭	2009	KIA	2.135

IV. 결론 및 제언

1. 결 론

본 연구는 한국 프로야구 타자의 경기력요인을 분석하기 위한 것으로 세이버메트릭스 지수를 활용하여 다양한 통계방법을 이용한 결과는 다음과 같다.

첫째, RC, XR, BR은 다른 세이버메트릭스 지수에 비해 상대적으로 편차가 크다. 이들의 계산공식은 매우 많은 타격지표를 필요로 하고 있어 값의 변화량이 큰 것으로 드러났다.

둘째, PSN과 BABIP는 다른 세이버메트릭스 지수들과 낮은 상관을 띄며, 군집간의 차이도 크지 않은 것으로 나타났다. 요인분석에서도 낮은 요인적재값을 보여 다른 지수들과의 관련성이 크지 않은 것으로 분석되었다. PSN은 호타준족을 나타내는 지수로 장타력을 보유하고 있지만 주력이 느린 타자와 주력이 높지만 장타력은 약한 타자들에게는 극히 낮은 수치를 나타낼 수밖에 없는 단점이 크게 작용하는 것으로 분석되며, BABIP는 타자의 타격스타일에 따라 개인의 고유한 BABIP 값을 가지게 되는데 예를 들어, 한화의 김태균이나 롯데의 손아섭 같은 경우 BABIP로 순위를 정했을 때 각각 1.3위와 2.4위를 기록할 정도로 BABIP 지수가 매우 높다. 이것은 라인드라이브성 타구를 많이 쳐내기 때문에 꾸준히 높은 BABIP 지수를 기록할 확률이 높다는 것을 의미한다. 그러나 MLB의 이치로의 경우를 보면 라인드라이브성 타구를 많이 치지 않더라도 높은 BABIP 지수를 가지는데 이는 이치로처럼 주력이 높은 선수는 땅볼타구를 치더라도 내야안타를 많이 만들어내기 때문에 이러한 유형의 선수 역시 높은 BABIP 지수를 꾸준히 기록할 가능성이 크다. 결국 BABIP 지수는 수년간 개인의 일정한 기록이 쌓인 경우 리그 평균의 BABIP와 비교하는 것보다 본인의 통산 BABIP와 비교하는 것이 훨씬 유의미한 것으로 분석된다.

셋째, 주성분회귀분석을 통해 산출된 요인점수 S에 표준화된 세이버메트릭스 지수를 대입하여 얻은 요인점수 순위에서 상위 10명의 가장 두드러진 특징으로는 년도에 상관없이 모두 홈런의 개수가 20홈런 이상이라는 점과 장타율이 많이 반영되는 OPS가 1,000 이상이라는 점으로 장타율이 선수들의 경기력요인에 많은 영향을 미치고 있음을 나타냈다.

2. 제 언

야구는 기록의 스포츠라는 말처럼 다양한 세이버메트릭스 지수를 활용하여 한국 프로야구 타자의 타격지표를 통해 경기력요인을 분석하면서 많은 한계점이 존재하고 있음을 확인하였다. 우선이영훈(2007), 이장택(2014a, 2014b), 장진희, 문춘걸(2014) 등이 언급한 바와 같이 국내에서 사용되고 있는 많은 세이버메트릭스 지수의 상수 값이 미국 MLB의 통산 기록에서 산출된 값이라는 점에서 이를 대체할만한 한국 프로야구 기록을 기반으로 하는 상수 값이 절실히 필요한 것으로 보인다. 미국 MLB와 한국 프로야구는 팀의 수와 선수의 수가 다르기 때문에 통산 타율과 득점, 타점 등의 기록에서 나타나는 수치의 차이를 배제하고 가져다 쓰는 것에는 무리가 따른다. 따라서 후속연구에서는 각각의 세이버메트릭스 지수에 대한 상수 값을 한국 프로야구 데이터를 토대로 산출할 것을 기대해본다.

최근 한국 프로야구에 대한 대중의 관심이 고조되면서 야구통계에 관한 지식도 더불어 각광을 받고 있다. 방송 매체를 통해서 다양한 통계수치들이 등장하고 있고, 이를 알기 쉽고 재미있게 이해시키려는 노력 또한 현재진행형이다. 따라서 보다 더 객관적이고 신뢰할만한 정보를 제공하기 위해서는 향후 많은 연구가 이루어져야 할 것이다.

● 참고 문헌 ●

- 강정구, 박석천, 김종현(2014). 세이버 메트릭스를 활용한 한국 프로야구 기록 시스템 제안. 한국인터넷정보학회 춘계학술발표대회 논문집, 15(1), 143-144.
- 박승현(2008). 한국프로야구 타자의 고액연봉에 영향을 미치는 경기력 요인. 한국체육과학회지, 17(2), 485-494.
- 승희배, 강기훈(2012). 한국 프로야구 선수들의 경기력과 연봉의 관계 분석. 한국데이터정보과학회지, 23(2), 285-298.
- 엄한주(2001). 요인분석 모형의 이해와 적용: 주성분모형과 공통요인모형의 방법론적 비교. 한국체육측정평가학회지, 3(1), 101-130.
- 원정주(2008). 세이버메트릭스 적용을 통한 테니스교육의 효율성 증대. 미간행 성균관대학교 석사학위논문.
- 이영훈(2007). 한국프로야구 경기력 결정요인에 관한 실증분석. 한국체육측정평가학회지, 9(2), 63-77.
- 이장영(2001). 프로야구 타자의 연봉과 경기능력의 관계. 한국사회체육학회지, 16, 973-981.
- 이장택(2014a). 한국프로야구에서 출루율 계수의 추정. 한국데이터정보과학회지, 25(2), 357-363.
- 이장택(2014b). 한국프로야구에서 타자능력의 측정. 한국데이터정보과학회지, 25(2), 349-356.
- 장진희, 문춘걸(2014). 한국 프로야구의 구단 승률에 대한 분석. 한국스포츠산업 · 경영학회지, 19(3), 17-31.
- 조영석, 조용주(2005). 한국 프로야구에서 OPS와 득점에 관한 연구. 한국자료분석학회지, 7(1), 221-231.
- 조영석, 조용주, 신상근(2007). 한국프로야구에서 승패 추정에 관한 연구. 한국자료분석학회지, 9(1), 501-510.
- 한국야구위원회(2009). 2009한국프로야구연감, 한국야구위원회, 서울.
- 한국야구위원회(2010). 2010한국프로야구연감, 한국야구위원회, 서울.
- 한국야구위원회(2011). 2011한국프로야구연감, 한국야구위원회, 서울.
- 한국야구위원회(2012). 2012한국프로야구연감, 한국야구위원회, 서울.
- 한국야구위원회(2013). 2013한국프로야구연감, 한국야구위원회, 서울.
- 한국야구위원회(2014). 2014한국프로야구연감, 한국야구위원회, 서울.
- Bill James(2001). The New Bill James Historical Baseball Abstract.
- Philip, B., Paul D. B., Bruce D. W.(2012). Predicting Run Production and Run Prevention in Baseball: The Impact of Sabermetrics. *International Journal of Business*, 2(4), 67-75.

ABSTRACT

Analysis of Professional Korean Baseball Batter's Performances Factors

Yang, Do-Eob · Cho, Eun-Hyung · Bae, Sang-Woo · Jung, Sang-Won

Modern baseball is a popular sport in many countries. Thus, numerous records are produced and accumulated data are analyzed by a variety of methods. The purpose of this study is to analyze performance of professional Korean baseball batters. The data were collected from official records issued by KBO(Korea Baseball Organization), and composed of the records of professional Korean Baseball 307 batters from 2008 to 2014. To analysis of the data was used 13 sabermetrics indexes. It's OPS, GPA, SECA, TA, RC, RC/27, XR, BR, ISO, PSN, wOBA, OW%, BABIP. As for its analysis method, descriptive statistics, correlation analysis, cluster analysis, principle component regression analysis were done by using SPSS Windows ver. 18.0.

RC, XR, BR showed a large deviation as compared with other indexes. PSN, BABIP has a low correlation with other indexes and no different between cluster. According to principle component regression analysis, slugging percentage showed a lot of influence on the performances of the batters.

Key Words : sabermetrics, professional baseball, K-means cluster analysis, principle component analysis

접 수 일 : 2015. 2. 28

논문심사일 : 2015. 4. 15

게재확정일 : 2015. 5. 29