

한국 프로야구에서 출루 능력과 장타력이 득점 생산성에 미치는 영향[†]

김혁주¹

¹원광대학교 수학·정보통계학부

접수 2012년 9월 13일, 수정 2012년 10월 4일, 게재확정 2012년 10월 22일

요약

한국 프로야구에서 출루 능력과 장타력이 득점 생산성에 미치는 영향을 통계적으로 분석하였다. 2절에서는 OPS (On-base percentage Plus Slugging average)에 관하여 고찰하고 그것을 변형한 새로운 타격지표들을 소개하였다. 3절에서는 2007년부터 2011년까지의 정규리그 전 경기 자료를 바탕으로 타율, 출루율, 장타율, IsoP (Isolated Power), OPS 및 2절에서 소개된 지표들이 팀의 경기당 평균득점과 갖는 상관관계를 알아보았다. 또한 OPS와 2절에서 소개된 지표들을 일반화하여 여러 가지 가중값을 갖는 타격지표들이 팀의 경기당 평균득점과 갖는 상관관계를 분석하였다. 그 결과 출루율에 57%, 장타율에 43%의 가중값을 주는 가중OPS가 득점 생산성을 가장 잘 설명하는 타격지표인 것으로 밝혀졌다.

주요용어: 가중오피에스, 득점, 상관계수, 순위상관계수, 오피에스, 한국 프로야구.

1. 서론

야구에서 팀이나 개인의 공격력을 나타내는 지표로 여러 가지가 있다. 그중에서 널리 알려진 것은 타율, 타점, 홈런, 출루율, 장타율 등이다. 이 중 출루율과 장타율을 정의하는 공식은 다음과 같다.

$$\text{출루율} = \frac{\text{안타} + 4\text{사구}}{\text{타수} + 4\text{사구} + \text{희생플라이}} \quad (1.1)$$

$$\text{장타율} = \frac{\text{단타} + 2 \times 2\text{루타} + 3 \times 3\text{루타} + 4 \times \text{홈런}}{\text{타수}} \quad (1.2)$$

위의 식에서 4사구란 4구 (base on balls)와 사구 (hit by pitched ball)를 통칭한 것이다. 간단히 말하면, 출루율이란 타석에 나와서 얼마나 자주 살아나갔느냐 하는 것으로서 야구 센스 (또는 정확도)를 나타내는 것이고, 장타율은 타수당 평균 몇 루타씩 쳤느냐 하는 것으로서 장타력을 나타내는 것이다.

타자들 중에는 출루율과 장타율이 모두 높은 타자도 있지만, 둘 중 한 가지만 높은 타자도 상당히 많다. 따라서 한 가지 척도로만 타자를 평가할 수는 없으며, 이는 타율, 타점, 홈런의 경우에도 마찬가지이다. 그리하여 몇 년 전부터는 OPS (On-base percentage Plus Slugging average)라는 지표가 사용되고 있다. 이것은

$$\text{OPS} = \text{출루율} + \text{장타율} \quad (1.3)$$

[†] 이 논문은 2010학년도 원광대학교의 교비 지원에 의해서 수행됨.

¹ (570-749) 전북 익산시 신용동 344-2, 원광대학교 수학·정보통계학부 및 기초자연과학연구소, 교수.
E-mail: hjkim@wonkwang.ac.kr

로 정의되는 지표로서, 정확성과 장타력을 종합하여 팀의 득점에 대한 기여도를 측정하기 위한 것이다.

다른 종목들도 마찬가지로이지만, 야구는 상대팀보다 득점을 많이 해야 이기는 경기이다. 따라서 타자들을 종합적으로 평가하는 타격지표도 팀의 득점에 얼마나 기여했는가를 기준으로 정할 필요가 있다. 이런 의미에서 OPS는 타자들의 타격 능력을 평가하는 공신력 있는 지표로 사용되고 있다.

그런데 OPS는 출루율과 장타율을 똑같은 비중으로 반영하고 있다는 점에서 야구 관계자들로부터 비판을 받고 있다. 출루율과 장타율의 단순합계인 OPS가 널리 쓰이는 이유는 단지 계산이 간편하다는 것이다. 그러나 출루율과 장타율 중 어느 한 쪽에 더 큰 비중을 둔 가중평균을 사용하여 팀의 득점 생산 능력을 좀 더 잘 설명할 수 있다면, 단순한 OPS보다 더욱 합리적인 타격지표가 될 것이다.

야구에서 득점과 타격 능력에 관한 국내 학자들의 대표적 연구는 다음과 같다. Cho와 Cho (2005a)는 한국 프로야구에서 OPS와 득점에 관한 연구를 통하여 OPS가 득점 추정에서 매우 좋은 결과를 준다는 점을 밝힌 바 있다. Lee와 Kim (2005)은 타자들의 공격 능력의 측정 도구로 제안되는 여러 가지의 득점 측도를 비교하였다. 한편 Cho와 Cho (2005b)는 한국 프로야구에서 득점과 실점을 이용한 승률 추정에 관하여 연구하였으며, Lee와 Lee (2008)는 한국 프로야구에서 팀의 승패를 예측하는 모형을 만들고, 어떤 요인들이 승패에 얼마만한 영향력을 미치는지 분석하였다. Hong과 Park (2008)은 한국 프로야구에서 팀들이 기록한 안타의 방향을 시각적으로 표현하고 통계적으로 분석하였으며, Chea 등 (2010)은 한국 프로야구의 정규리그 자료를 이용하여 포스트시즌 진출 여부를 예측할 수 있는 여러 가지 통계모형의 분석 결과를 비교하였다. 그리고 야구 관중 수에 관한 연구로서, Lee와 Bang (2010)은 일반화 자기회귀 조건부 이분산 모형을 이용하여 한국 프로야구 관중 수를 예측하는 방법을 제안하였다.

2. OPS와 그의 두 가지 변형

2.1. OPS

1절에서 언급한 바와 같이 OPS는 식 (1.3)으로 정의되는 지표이다. 출루율과 장타율의 단순합으로 정의되는 OPS를 사용하는 것이 득점 생산성을 설명하는 데 최적인가 하는 것은 3절에서 여러 가지의 가중값을 사용한 가중OPS를 살펴봄으로써 규명하겠지만, 여기서는 우선 OPS의 다음과 같은 측면을 주목한다. (여기서 말하는 성질은 일반적인 것이다. 아주 특수한 예외로, 예컨대 시즌의 첫 타석에서 2루타를 친 타자가 다음 타석에서 단타를 치면 OPS가 오히려 감소한다. 이것은 그 다음에 소개하는 지표A와 지표B의 경우도 마찬가지이다.)

- (1) 단타를 치는 경우: 출루율과 장타율이 동시에 증가한다. 따라서 단타에 의한 OPS의 증가에는 중복되는 면이 있다.
- (2) 장타를 치는 경우: 역시 출루율과 장타율이 동시에 증가한다. 2루타, 3루타, 홈런 중 1루타에 해당하는 장타율 증가분은 출루율의 증가와 중복되는 면이 있고, 나머지 증가분이 장타에 의한 추가 증가분이다.
- (3) 4사구를 얻는 경우: 출루율만 증가하고 장타율에는 변함이 없다.

즉 안타 (단타 또는 장타)를 쳤을 때 OPS의 증가분에는 출루율의 증가와 장타율의 증가가 모두 반영된다. 따라서 득점 생산성과의 관련성 연구에서 다음과 같은 새로운 지표들 (중복성이 없는)이 더욱 적절할지도 모른다는 추측이 들어 OPS와 함께 연구하기로 한다.

2.2. 지표A

이것은 다음 식으로 정의되는 타격지표이다.

$$\text{지표A} = \text{출루율} + \text{IsoP} \quad (2.1)$$

여기서 IsoP (Isolated Power)란

$$\text{IsoP} = \text{장타율} - \text{타율} \quad (2.2)$$

로 정의되는 지표이다. 식 (1.2)의 장타율이 단타만 쳐도 (심지어 내야안타만 쳐도) 증가하는 지표이므로 순수한 장타력을 나타내지 못한다는 개념에서 제안된 이 IsoP는 절대장타율이라고도 불린다. 따라서 장타율 대신 IsoP를 출루율에 더하여 새로운 지표를 생각하고 이것을 OPS와 비교해 보는 것은 자연스러운 시도라고 할 수 있다. 이 지표를 편의상 지표A라 하자.

2.3. 지표B

이것은 다음 식으로 정의되는 타격지표이다.

$$\text{지표B} = \frac{\text{4사구}}{\text{타수} + \text{4사구} + \text{희생플라이}} + \text{장타율} \quad (2.3)$$

지표A가 안타를 쳤을 때 OPS에 내재된 중복성을 장타율 쪽에서 조정한 것인 반면, 지표B는 이 중복성을 출루율 쪽에서 조정한 것이다.

3. 타격지표들과 득점 생산성 사이의 관계 분석

3.1. OPS, 지표A, 지표B

우선 여러 가지의 타격지표와 득점 생산성 사이의 관계를 살펴보자. Table 3.1에는 Korea Baseball Organization (2008, 2009, 2010, 2011, 2012)의 2007, 2008, 2009, 2010, 2011 Official Baseball Guide에서 자료를 구할 수 있는 최근 5년간 (2007년~2011년)의 각 팀의 정규리그 전 경기 타격 자료가 나와 있다. 이 표에서 영문 약자는 각각 다음을 나타내며, Table 3.2에서도 마찬가지이다.

Year: 연도, Team: 팀, Game: 경기 수, Run: 득점, RPG: 경기당 평균득점

BA: 타율, OBP: 출루율, Slg: 장타율, Index A: 지표A, Index B: 지표B

이 자료를 바탕으로 하여 타율과 출루율을 포함한 여러 지표와 경기당 평균득점 간의 상관계수를 구하였다. (본 논문의 모든 통계분석과 그래프 작성은 Minitab Release 14.12.1을 이용하여 실시하였다.) 또한 참고로 각각의 지표를 기준으로 하여 팀들의 순위를 구한 뒤 경기당 평균득점을 기준으로 한 순위와의 순위상관계수를 구하여 Table 3.2에 함께 나타냈다. 이 표에서 볼 수 있듯이 상관계수와 순위상관계수는 모든 지표에서 매우 유사한 값을 갖는다. 여러 지표들을 비교해 보면, 타율이 평균득점과 가장 낮은 상관관계를 보이고 있다. 따라서 득점 생산에 대한 기여도를 논할 때 타율은 배제하고 출루율과 장타율에 주목하는 것이 타당하며, 출루율과 장타율을 결합한 OPS 및 2.2절과 2.3절에서 고려한 지표A와 지표B에 초점을 맞추는 것이 합리적이다. Table 3.2를 보면, 여러 지표 중 OPS가 평균득점과 가장 높은 양의 상관관계 (상관계수 0.94019)를 보이고 있으므로, 현재 사용되고 있는 타격지표 중에서는 OPS가 팀의 득점 생산성을 설명하는 가장 적절한 지표이다. OPS는 지표A와 지표B보다도 강한 설명력을 보인다.

Table 3.1 Batting data of the whole regular season (2007~2011)

Year	Team	Game	Run	RPG	BA	OBP	Slg	IsoP	OPS	Index_A	Index_B
2007	SK	126	603	4.785714	0.263928	0.341260	0.402738	0.138810	0.743998	0.480070	0.510790
2007	Doosan	126	578	4.587302	0.262682	0.342595	0.383425	0.120743	0.726019	0.463338	0.495603
2007	Hanwha	126	534	4.238095	0.254360	0.341670	0.376211	0.121851	0.717881	0.463521	0.495330
2007	Samsung	126	497	3.944444	0.253703	0.340904	0.365265	0.111562	0.706170	0.452467	0.484873
2007	LG	126	532	4.222222	0.268394	0.338357	0.374364	0.105969	0.712720	0.444326	0.473249
2007	Hyundai	126	530	4.206349	0.270892	0.345568	0.383568	0.112676	0.729136	0.458244	0.488469
2007	Lotte	126	533	4.230159	0.269600	0.340292	0.376550	0.106951	0.716843	0.447243	0.476342
2007	Kia	126	499	3.960317	0.256898	0.330304	0.363701	0.106803	0.694006	0.437107	0.465772
2008	SK	126	632	5.015873	0.282087	0.360553	0.403970	0.121884	0.764523	0.482436	0.516300
2008	Doosan	126	647	5.134921	0.275783	0.353989	0.377993	0.102210	0.731982	0.456199	0.489374
2008	Lotte	126	624	4.952381	0.281955	0.356462	0.404605	0.122650	0.761068	0.479113	0.512517
2008	Samsung	126	557	4.420635	0.257895	0.344194	0.376077	0.118182	0.720270	0.462376	0.495124
2008	Hanwha	126	592	4.698413	0.254035	0.332505	0.395322	0.141287	0.727827	0.473792	0.502982
2008	Kia	126	503	3.992063	0.260095	0.336283	0.352257	0.092162	0.688540	0.428445	0.457819
2008	Heroes	126	499	3.960317	0.266157	0.330755	0.367616	0.101459	0.698371	0.432214	0.458145
2008	LG	126	468	3.714286	0.255736	0.320593	0.353250	0.097514	0.673843	0.418107	0.444081
2009	Kia	133	706	5.308271	0.267340	0.357129	0.427385	0.160045	0.784514	0.517174	0.552847
2009	SK	133	732	5.503759	0.284871	0.369333	0.452558	0.167687	0.821891	0.537020	0.573701
2009	Doosan	133	722	5.428571	0.280113	0.361519	0.419734	0.139621	0.781254	0.501140	0.535787
2009	Lotte	133	637	4.789474	0.276697	0.349890	0.422624	0.145928	0.772514	0.495817	0.527571
2009	Samsung	133	685	5.150376	0.274501	0.363866	0.431711	0.157210	0.795577	0.521076	0.558045
2009	Heroes	133	683	5.135338	0.271734	0.355560	0.433794	0.162060	0.789354	0.517620	0.552379
2009	LG	133	670	5.037594	0.277691	0.359542	0.413319	0.135627	0.772860	0.495169	0.529995
2009	Hanwha	133	657	4.939850	0.268615	0.347992	0.426407	0.157792	0.774399	0.505785	0.537114
2010	SK	133	704	5.293233	0.274161	0.355310	0.412480	0.138319	0.767790	0.493629	0.527561
2010	Samsung	133	681	5.120301	0.272092	0.367658	0.405996	0.133904	0.773655	0.501563	0.540463
2010	Doosan	133	731	5.496241	0.281181	0.365140	0.440062	0.158881	0.805202	0.524021	0.559723
2010	Lotte	133	773	5.812030	0.287824	0.352240	0.460518	0.172694	0.812758	0.524934	0.554154
2010	Kia	133	611	4.593985	0.259964	0.347241	0.382699	0.122736	0.729941	0.469977	0.502929
2010	LG	133	688	5.172932	0.275847	0.348882	0.410513	0.134666	0.759395	0.483548	0.514599
2010	Nexen	133	570	4.285714	0.261680	0.339679	0.373989	0.112309	0.713668	0.451988	0.482655
2010	Hanwha	133	543	4.082707	0.243694	0.330562	0.358945	0.115251	0.689506	0.445813	0.475959
2011	Samsung	133	625	4.699248	0.259184	0.343325	0.375930	0.116746	0.719254	0.460070	0.492475
2011	SK	133	584	4.390977	0.262716	0.340366	0.385332	0.122616	0.725697	0.462981	0.493228
2011	Lotte	133	713	5.360902	0.287889	0.358399	0.422483	0.134594	0.780883	0.492994	0.524938
2011	Kia	133	627	4.714286	0.269152	0.356574	0.395381	0.126229	0.751954	0.482803	0.517492
2011	Doosan	133	614	4.616541	0.271454	0.353732	0.388395	0.116941	0.742127	0.470673	0.504162
2011	LG	133	579	4.353383	0.266033	0.338198	0.377488	0.111455	0.715685	0.449653	0.478454
2011	Hanwha	133	568	4.270677	0.254637	0.337909	0.361575	0.106938	0.699484	0.444847	0.475019
2011	Nexen	133	512	3.849624	0.245222	0.322554	0.353442	0.108220	0.675997	0.430774	0.458027

Table 3.2 Correlation coefficient and rank correlation coefficient between each index and average runs per game

	BA	OBP	Slg	IsoP	OPS	Index A	Index B
Corr. coeff.	0.80386	0.85861	0.91649	0.82398	0.94019	0.91197	0.91052
Rank corr. coeff.	0.79610	0.86702	0.91074	0.82696	0.94160	0.90820	0.91121

3.2. 가중OPS

현재 사용되고 있는 여러 지표 중 득점 생산성을 가장 잘 나타내고 있는 것으로 밝혀진 OPS는 출루율과 장타율의 단순합계이다. 이것은 출루율과 장타율의 비중을 똑같이 50%씩 둔 것이다. 그러나 이 비중을 일반화하여 가능한 경우를 모두 살펴본다면, 팀의 득점 생산성을 더욱 잘 설명하는 타격지표를 찾아낼 수 있을 것으로 기대한다. 이런 뜻에서 다음의 식으로 표시되는 가중OPS를 정의한다.

$$\text{가중OPS} = a \times \text{출루율} + (1 - a) \times \text{장타율} \tag{3.1}$$

여기서 a 는 0과 1 사이의 값으로서 출루율의 가중값을 나타낸다. $a = 0$ 인 경우에 가중OPS는 장타율이 되고, $a = 1$ 인 경우에는 출루율이 되며, $a = 0.5$ 인 경우에는 기존의 OPS와 같은 개념을 갖는 것이 된다. a 의 값을 0부터 1까지 0.01 간격으로 변화시켜 가며 팀들의 가중OPS를 구하고, 이 가중OPS와 팀의 경기당 평균득점 간의 상관계수와 순위상관계수를 구하여, a 값에 따라 변화하는 내용을 그래프

로 작성한 것이 Figure 3.1과 같다. 이 그림에서 실선은 상관계수를 나타내고 점선은 순위상관계수를 나타낸다.

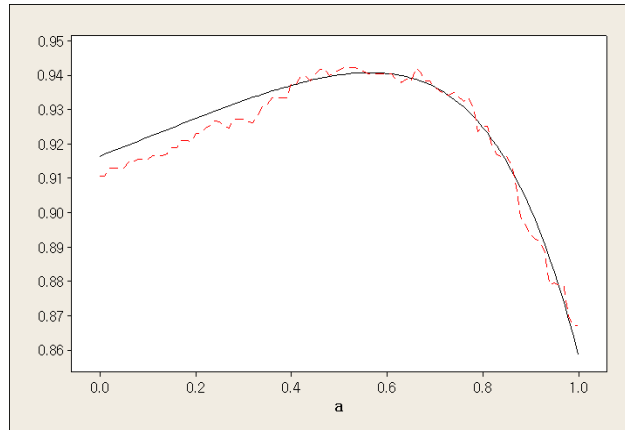


Figure 3.1 Plot of correlation coefficient and rank correlation coefficient against a (weighted OPS)

상관계수는 $a = 0.57$ 일 때 최댓값 0.94087을 갖는 것으로 밝혀졌다. 즉 가중OPS 중에서는

$$\text{가중OPS}(0.57) = 0.57 \times \text{출루율} + 0.43 \times \text{장타율} \quad (3.2)$$

로 정의되는 지표를 사용할 때 팀들의 득점 생산성을 가장 잘 설명할 수 있었다. 순위상관계수는 $a = 0.51, 0.52, 0.53$ 일 때 최댓값 0.94235를 갖는 것으로 밝혀졌고, $a = 0.57$ 일 때는 0.94048로서 최댓값과 거의 차이가 없었다. 상관계수를 최대로 하는 a 값과 순위상관계수를 최대로 하는 a 값에 약간 차이가 있으나, 지표의 값들로부터 직접 계산된 상관계수가 우선이므로 식 (3.2)로 정의된 지표를 사용하는 것이 타당한 것으로 사료된다.

3.3. 가중지표A

OPS에서 출루율과 장타율에 여러 가지의 가중값을 줘서 득점 생산성을 가장 잘 설명하는 가중OPS를 찾아낸 것과 같이 지표A에서도 출루율과 IsoP에 여러 가중값을 줘서 다음과 같은 가중지표A를 고려한다.

$$\text{가중지표A} = a \times \text{출루율} + (1 - a) \times \text{IsoP} \quad (3.3)$$

a 의 값을 0부터 1까지 0.01 간격으로 변화시켜 가며 팀들의 가중지표A를 구하고, 이 가중지표A와 팀의 경기당 평균득점 간의 상관계수와 순위상관계수를 구한 결과 상관계수는 $a = 0.68$ 일 때 최댓값 0.92452를 가지며, 순위상관계수는 $a = 0.63$ 일 때 최댓값 0.93344를 갖는 것으로 밝혀졌다.

3.4. 가중지표B

지표B에서도 역시 첫 항과 둘째 항에 여러 가중값을 줘서 다음과 같은 가중지표B를 고려한다.

$$\text{가중지표B} = a \times \frac{\text{4사구}}{\text{타수} + \text{4사구} + \text{희생플라}} + (1 - a) \times \text{장타율} \quad (3.4)$$

a 의 값을 0부터 1까지 0.01 간격으로 변화시켜 가며 팀들의 가중지표B를 구하고, 이 가중지표B와 팀의 경기당 평균득점 간의 상관계수와 순위상관계수를 구한 결과 상관계수는 $a = 0.28$ 일 때 최댓값 0.92354를 가지며, 순위상관계수는 $a = 0.18$ 일 때 최댓값 0.92199를 갖는 것으로 밝혀졌다.

3.5. 논의

3.1절부터 3.4절까지의 내용을 종합하면, 팀의 득점 생산성을 가장 잘 설명해 주는 타격지표는 식 (3.2)로 표시된 가중OPS(0.57)인 것으로 나타났으며, 팀의 경기당 평균득점과의 상관계수는 0.94087로 매우 높았다. 출루율의 상관계수 (0.85861)보다 장타율의 상관계수 (0.91649)가 더 높았지만 최적의 가중OPS에서는 출루율에 더 큰 가중값이 붙는다는 점이 주목할 만하다. 가중OPS(0.57)과 팀의 경기당 평균득점 간의 산점도를 최소제곱회귀직선과 함께 나타낸 것이 Figure 3.2이다. 가중OPS(0.57)을 x , 평균득점을 y 라 할 때, 최소제곱회귀직선의 방정식은

$$\hat{y} = -5.546 + 27.84x$$

였다. 결정계수는 88.5%로 매우 높았으며, 회귀분석의 F 통계량의 값은 293.12였고 p 값은 0.001보다 작았다.

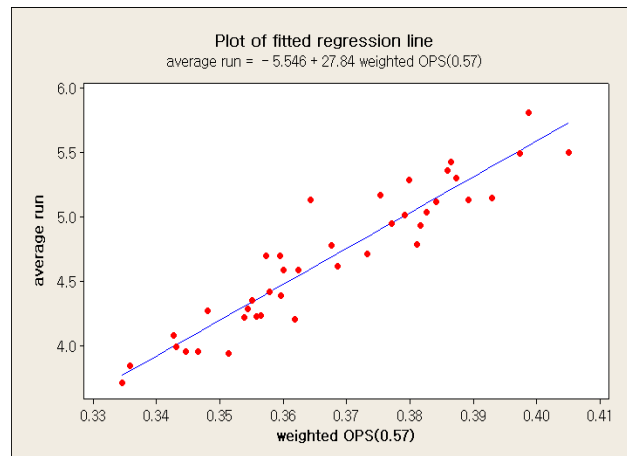


Figure 3.2 Plot of average run against weighted OPS with $a=0.57$ (2007~2011)

2007년부터 2011년까지 연도별로 가중OPS(0.57)을 지표로 하여 규정타석을 채운 타자들의 순위를 1위부터 10위까지 나타낸 것이 Table 3.3과 같다. (표에서 W_OPS가 가중OPS(0.57)을 나타낸다.) 이 순위를 팀의 득점에 기여한 정도로 해석할 수 있을 것이다. 기존의 OPS를 기준으로 한 순위 (OPS.R)도 함께 표시하였다. 두 지표에 의한 결과에 약간의 차이가 있는 것을 볼 수 있다. 특히 2008년, 2009년, 2010년에는 5위권 이내에서도 순위의 변동이 생겼다.

Table 3.3 Top 10 batters based on weighted OPS(0.57) for each year 2007

Rank	Batter	Team	W_OPS	OPS_R
1	LeeDH	Lotte	0.516	1
2	YangJH	Samsung	0.502	2
3	KimDJ	Doosan	0.490	3
4	Brumbaugh	Hyundai	0.480	4
5	Cruz	Hanwha	0.477	5
6	KimTK	Hanwha	0.447	6
7	SimJS	Samsung	0.438	7
8	JeongKW	SK	0.433	8
9	ParkJM	Samsung	0.4260	10
10	LeeHJ	SK	0.4255	9

2008

Rank	Batter	Team	W_OPS	OPS_R
1	KimTK	Hanwha	0.505	1
2	KimHS	Doosan	0.478	2
3	ParkJH	SK	0.471	3
4	KimDJ	Doosan	0.452	4
5	ChoiJ	SK	0.441	6
6	Garcia	Lotte	0.4339	5
7	LeeDH	Lotte	0.4338	7
8	LeeTK	Woori	0.429	8
9	ChoSH	Lotte	0.4169	12
10	KangMH	Lotte	0.4168	11

2009

Rank	Batter	Team	W_OPS	OPS_R
1	KimDJ	Doosan	0.515	1
2	Petagine	LG	0.514	2
3	KimHS	Doosan	0.508	3
4	ChoiHS	Kia	0.501	4
5	ParkYT	LG	0.4882	6
6	KimSH	Kia	0.4877	5
7	HongSH	Lotte	0.477	7
8	KimTW	Hanwha	0.471	8
9	JeongKW	SK	0.457	9
10	KangBK	Samsung	0.448	10

2010

Rank	Batter	Team	W_OPS	OPS_R
1	LeeDH	Lotte	0.540	1
2	HongSH	Lotte	0.502	2
3	KimHS	Doosan	0.464	3
4	ChoiJS	Doosan	0.460	4
5	ParkSM	Samsung	0.457	7
6	ChoiJ	SK	0.455	6
7	ChoIS	LG	0.453	5
8	KimDJ	Doosan	0.450	8
9	ChoiHW	Samsung	0.447	9
10	KangMH	Lotte	0.440	10

2011				
Rank	Batter	Team	W OPS	OPS_R
1	ChoiHW	Samsung	0.509	1
2	LeeDH	Lotte	0.496	2
3	ChoiJ	SK	0.455	3
4	SonAS	Lotte	0.437	4
5	KimDJ	Doosan	0.429	5
6	LeeBK	LG	0.423	6
7	LeeYK	Kia	0.416	9
8	KangMH	Lotte	0.4148	7
9	KimHS	Doosan	0.4145	8
10	ChoiJH	Hanwha	0.405	10

4. 결론

팀과 개인의 타격 능력을 나타내는 지표 중 두 축을 이루는 것이 출루율과 장타율이다. 득점 생산성과 관련하여 생각할 때는 특히 그렇다. 타격지표 중 타율이 가장 간단하고 기본적인 것이기는 하지만, 이것은 4사구를 고려하지 않은 것이므로, 득점과 관련해서 생각할 때는 중요성에서 출루율보다 못하며, 루타수를 고려하지 않은 것이므로 장타율보다도 못하다. 3.1절의 결과도 이를 잘 보여주고 있다.

이런 의미에서 출루율과 장타율을 결합한 OPS가 득점 생산성과 관련된 타격지표로 널리 사용되고 있다. 2007년부터 2011년까지의 한국 프로야구 정규리그 자료를 근거로 할 때, Table 3.2에서 보는 바와 같이 OPS는 함께 고려된 다른 지표들보다 팀 득점과의 관련성이 높았다. 본 논문에서는 안타를 치는 경우 OPS의 증가에 내재되는 중복성에 주목하여 지표A와 지표B라는 지표를 설정하고 득점과의 관련성을 살펴보았으나, OPS가 이 지표들보다도 더 큰 설명력을 갖는 것으로 밝혀졌다. 이어서 OPS와 지표A 및 지표B에서 가중값을 여러 가지로 바꿔 가며 팀 득점과의 관계를 분석한 결과, 출루율에 57%의 비중을 두고 장타율에 43%의 비중을 두는 가중OPS가 팀 득점 생산성에 대하여 가장 높은 설명력을 보였다. 이는 출루율에 좀 더 큰 가중값을 줘서 팀이나 타자의 타격 능력을 평가하는 것이 타당함을 말해 준다.

본 논문에서는 최근 5년간의 자료를 근거로 연구했으나, 자료 접근이 가능하다면 후속 연구로 프로야구 원년부터의 자료를 사용하여 시대 변화에 따른 추이를 연구하는 것도 의미가 있을 것이다. 시대가 바뀐다면 득점 생산성을 가장 잘 설명하는 가중OPS에서 출루율과 장타율의 가중값이 본 논문에서 얻어진 것과 약간 달라질 수도 있기 때문이다. 그밖에도 도루 등 다른 요인들을 고려한 연구나 팀별 연구, 또는 득점을 저지하기 위한 상대팀의 수비력 측면에서의 연구도 가능할 것으로 전망된다.

참고문헌

- Chea, J. S. Cho, E. H. and Eom, H. J. (2010). Comparisons of the outcomes of statistical models applied to the prediction of post-season entry in Korean professional baseball. *The Korean Journal of Measurement and Evaluation in Physical Education and Sport Science*, **12**, 33-48.
- Cho, Y.-S. and Cho, Y.-J. (2005a). A study on OPS and runs from Korean baseball league. *Journal of the Korean Data Analysis Society*, **7**, 221-231.
- Cho, Y.-S. and Cho, Y.-J. (2005b). A study on winning percentage using batter's runs and pitcher's runs in Korean professional baseball league. *Journal of the Korean Data Analysis Society*, **7**, 2303-2312.
- Hong, C. S. and Park, H. (2008). Visual representation and applications of hitting direction in Korean baseball records. *Journal of the Korean Data & Information Science Society*, **19**, 539-549.

- Korea Baseball Organization (2008). *2007 Official baseball guide*, Korea Baseball Organization, Seoul.
- Korea Baseball Organization (2009). *2008 Official baseball guide*, Korea Baseball Organization, Seoul.
- Korea Baseball Organization (2010). *2009 Official baseball guide*, Korea Baseball Organization, Seoul.
- Korea Baseball Organization (2011). *2010 Official baseball guide*, Korea Baseball Organization, Seoul.
- Korea Baseball Organization (2012). *2011 Official baseball guide*, Korea Baseball Organization, Seoul.
- Lee, H.-Y. and Lee, S.-K. (2008). Relation analysis between victory and the records of Korean professional baseball. *Journal of the Korean Data Analysis Society*, **10**, 3413-3422.
- Lee, J. and Bang, S. (2010). Forecasting attendance in the Korean professional baseball league using GARCH models. *Journal of the Korean Data & Information Science Society*, **21**, 1041-1049.
- Lee, J.-T. and Kim, Y.-T. (2005). A study on runs evaluation measure for Korean pro-baseball players. *Journal of the Korean Data Analysis Society*, **7**, 2289-2302.

Effects of on-base and slugging ability on run productivity in Korean professional baseball[†]

Hyuk Joo Kim¹

¹Division of Mathematics & Informational Statistics and Institute of Basic Natural Sciences,
Wonkwang University

Received 13 September 2012, revised 4 October 2012, accepted 22 October 2012

Abstract

The purpose of this paper is to statistically analyze the effects of on-base and slugging ability on the run productivity in Korean professional baseball. In Section 2, we have investigated the OPS (On-base percentage Plus Slugging average) and introduced new indices of batting ability by modifying the OPS. In Section 3, we have examined the correlation which the batting average, on-base percentage, slugging average, IsoP (Isolated Power), OPS and the indices introduced in Section 2 have with the average runs per game, using the data from all the games of the regular seasons in 2007~2011. In addition, by generalizing the OPS and the indices introduced in Section 2, we have analyzed the correlation of the indices with various weights between the average runs per game. As a result, the weighted OPS consisting of on-base percentage (with weight 57%) and slugging average (with weight 43%) has been found to give the best explanation of the run productivity.

Keywords: Correlation coefficient, Korean professional baseball, OPS, rank correlation coefficient, run, weighted OPS.

[†] This paper was supported by Wonkwang University in 2010.

¹ Professor, Division of Mathematics & Informational Statistics and Institute of Basic Natural Sciences, Wonkwang University, Iksan 570-749, Korea. E-mail: hjkim@wonkwang.ac.kr