

Explanation of Run Productivity Using Weighted Adjusted OPS in Korean Professional Baseball

Hyuk Joo Kim^{a,1} · Yea Hyoung Kim^a

^aDivision of Mathematics & Informational Statistics, Wonkwang University

(Received July 21, 2014; Revised September 18, 2014; Accepted October 2, 2014)

Abstract

We suggested an adjusted OPS and weighted adjusted OPS as indices to explain run productivity of teams using the slugging average and adjusted OBP from Korean professional baseball. First, we defined adjusted OBP by modifying currently used OBP. Next, we defined adjusted OPS as the sum of adjusted OBP and slugging average. We also defined weighted adjusted OPS as the weighted average of adjusted OBP and slugging average. Analysis of the data from all games in the regular seasons from 1982 ~ 2013 shows that adjusted OPS better explains runs than OPS. For 25 seasons out of 32 seasons, adjusted OPS explains runs better than OPS. Further, weighted adjusted OPS consisting of adjusted OBP (with weight 60%) and slugging average (with weight 40%) gives the best explanation of run productivity. Weighted adjusted OPS has been found to explain run productivity better than weighted OPS proposed in Kim (2012).

Keywords: Adjusted OBP, adjusted OPS, weighted adjusted OPS, run, correlation coefficient.

1. 서론

세이버메트릭스(Sabermetrics)가 발전하면서 야구의 여러 가지 측면이 통계적으로 연구되고 있다. 팀의 득점과 타격 능력에 관한 국내 학자들의 연구도 활발한데, 최근 몇 년 간의 대표적 연구를 보면 다음과 같다. Cho와 Cho (2005a)는 한국 프로야구의 자료를 바탕으로 한 OPS(On-base percentage Plus Slugging average)와 득점에 관한 연구를 통하여 OPS가 득점 추정에서 매우 좋은 결과를 준다는 점을 밝혔다. Lee와 Kim (2005)은 타자들의 공격 능력의 측정 도구로 제안되는 여러 가지의 득점 측도를 비교하였다. 한편 Cho와 Cho (2005b)는 한국 프로야구에서 득점과 실점을 이용하여 승률을 추정하는 문제를 연구하였으며, Lee와 Lee (2008)는 한국 프로야구에서 팀의 승패를 예측하는 모형을 만들고, 각종의 요인들이 승패에 미치는 영향력을 분석하였다. Hong과 Park (2008)은 한국 프로야구에서 팀들이 기록한 안타의 방향을 분석하고 이를 시각적으로 표현하는 방법을 개발하였으며, Chea 등 (2010)은 한국 프로야구의 정규리그 자료를 이용하여 포스트시즌 진출 여부를 예측할 수 있는 여러 가지 통계모형의 분석 결과를 비교하였다. 또한 Lee (2014a)는 한국 프로야구에서 타자의 능력을 측정하는 지표를 제안하였으며, Lee (2014b)는 한국 프로야구에서 출루율 계수를 추정하는 문제를 연구하였다.

This paper was supported by Wonkwang University in 2012.

¹Corresponding author: Division of Mathematics & Informational Statistics and Institute of Basic Natural Sciences, Wonkwang University, 460, Iksan-daero, Iksan, Jeonbuk 570-749, Korea.
E-mail: hjkim@wonkwang.ac.kr

야구에서 팀이나 개인의 공격력을 나타내는 지표에는 여러 가지가 있다. 그중에서 최근 몇 년 사이에 타자의 능력을 평가하는 중요한 지표로 널리 사용되고 있는 것이 OPS인데, OPS는 다음 식과 같이 정의된다.

$$\text{OPS} = \text{출루율} + \text{장타율}, \quad (1.1)$$

여기서 출루율과 장타율은 다음과 같이 정의되는 지표들이다.

$$\text{출루율} = \frac{\text{안타수} + 4\text{사구수}}{\text{타수} + 4\text{사구수} + \text{희생플라이수}}, \quad (1.2)$$

$$\text{장타율} = \frac{\text{단타수} + 2 \times 2\text{루타수} + 3 \times 3\text{루타수} + 4 \times \text{홈런수}}{\text{타수}}. \quad (1.3)$$

위의 식에서 4사구란 4구(base on balls)와 사구(hit by pitched ball)를 통칭한 것이다.

간단히 말하면, 출루율이란 정확성(살아나간 비율)을 나타내는 것이고, 장타율은 장타력(타수당 평균 루타수)을 나타내는 것이다. 이 정확성과 장타력은 타격 능력을 평가하는 두 개의 축이라 할 수 있는데, 타자들 중에는 이 두 가지를 균형 있게 갖춘 타자도 있지만, 한 쪽으로 치우친 타자들도 상당히 많다. 이러한 의미에서 OPS는 정확성과 장타력을 종합하여 측정하기 위한 것이라고 할 수 있으며, 타격 능력을 평가하는 공신력 있는 지표로 사용되고 있다.

한편 Kim (2012)은, 현재 사용되고 있는 OPS가 출루율과 장타율을 똑같은 비중으로 반영하고 있다는 점에 주목하여 출루율과 장타율 중 어느 한 쪽에 더 큰 비중을 둔 가중평균인 가중OPS를 제시하였다. 또한 이를 2007년부터 2011년까지의 한국 프로야구에 적용한 결과 출루율에 57%, 장타율에 43%의 비중을 두는 가중OPS가 팀의 득점력을 가장 잘 설명할 수 있음을 보였다.

그런데 이 가중OPS를 좀 더 개선하면 팀의 득점 생산 능력을 더욱 잘 설명하는 공격지표를 만들어낼 수 있다. 본 논문에서는 출루율을 수정한 수정출루율과 이를 바탕으로 한 수정OPS를 사용할 것을 제시하고 수정출루율과 장타율의 가중평균으로 정의되는 가중수정OPS를 제시하여 팀의 득점력에 대한 설명력을 더욱 향상시키고자 한다.

2. 수정출루율과 수정OPS

본 논문에서는 다음 식으로 정의되는 수정출루율과 수정OPS를 사용할 것을 제안한다.

$$\text{수정출루율} = \frac{\text{안타수} + 4\text{사구수}}{\text{타수} + 4\text{사구수}}, \quad (2.1)$$

$$\text{수정OPS} = \text{수정출루율} + \text{장타율}. \quad (2.2)$$

즉 수정출루율은 현재의 출루율의 분모에서 희생플라이수를 제거한 것이며, 수정OPS는 수정출루율과 장타율을 합친 것이다. 한국 프로야구에서 1982년부터 1985년까지는 출루율을 식 (2.1)로 정의하여 사용하였으나, 1986년부터는 식 (1.2)로 정의하여 사용하고 있다.

팀의 득점에 대한 설명과 관련하여 식 (2.1)로 정의되는 수정출루율을 사용할 것을 제안하게 된 동기는 다음과 같다. 식 (1.2)에서 볼 수 있듯이 현재의 출루율에서는 타자가 희생플라이를 기록하면 출루율이 낮아진다. 그러나 타율(= 안타수/타수)을 계산할 때는 희생플라이수는 분모에 포함되지 않는다. 희생플라이는 타자가 희생해서 같은 팀의 주자를 진루시키거나 홈으로 들어오게 한 것이므로 타율에서 불이익을 보지 않게 하려는 것이다. 이처럼 타율과 출루율에서 희생플라이를 처리하는 관점이 다르다. 따라

Table 3.1. Correlation coefficients between OPS, AOPS and average runs per game

Year	1982	1983	1984	1985	1986	1987	1988	1989
OPS	0.88426	0.82629	0.99170	0.96352	0.95450	0.95804	0.83071	0.95237
AOPS	0.88253	0.83044	0.99354	0.96392	0.95558	0.95862	0.83137	0.95535
Year	1990	1991	1992	1993	1994	1995	1996	1997
OPS	0.95425	0.97077	0.95285	0.99026	0.96884	0.84857	0.88887	0.96939
AOPS	0.95524	0.97201	0.95353	0.99036	0.96841	0.85581	0.88661	0.96898
Year	1998	1999	2000	2001	2002	2003	2004	2005
OPS	0.94001	0.95265	0.94827	0.87709	0.96814	0.98876	0.95800	0.70704
AOPS	0.94097	0.95285	0.94858	0.88292	0.96895	0.98834	0.95745	0.71649
Year	2006	2007	2008	2009	2010	2011	2012	2013
OPS	0.90813	0.88019	0.91993	0.70645	0.98088	0.94389	0.87695	0.95961
AOPS	0.91327	0.89144	0.92020	0.70164	0.98111	0.94511	0.89076	0.96132

서 본 논문에서는 타율에서 희생플라이를 처리하는 방식을 출루율에도 적용해서, 희생플라이를 친 타자의 출루율을 낮아지지 않게 하려는 것이다. 이 수정출루율에 장타율을 더한 것이 수정OPS이다.

이와 같이 수정OPS를 사용하는 것은 OPS를 정의할 때 타자의 기여도와 가치를 평가하는 기본 개념과도 좀 더 일치한다고 할 수 있다. 야구에서 공격의 기본 메커니즘은 출루와 진루이다. 전자는 출루율에 의해 측정되고 후자는 장타율에 의해 측정된다(도루도 있지만, 장타율의 비중이 훨씬 크다). 즉 장타를 친 타자는 자기 자신이 많이 진루를 할 뿐 아니라, 주자가 있는 경우 주자도 많이 진루시키는 것이다. 그러나 희생플라이는 (이것이야말로 주자를 확실히 진루시키는 것임에도 불구하고) 장타율에 전혀 반영되지 않으며, 출루율에서는 오히려 값을 떨어지게 한다. 따라서 식 (2.1)로 정의되는 수정출루율과 이에 따른 수정OPS를 사용하면 희생플라이의 기여도를 더욱 정확하게 평가하게 되어 결국 팀의 득점을 더욱 잘 설명하는 지표가 될 것으로 기대한다.

3. 타격지표들과 득점력 사이의 관계 분석

본 논문에서 사용된 자료는 한국 프로야구 원년인 1982년부터 2013년까지 32년에 걸친 정규리그 전 경기(15,084 경기)의 타격자료이다. 이 자료는 Korea Baseball Organization (2009)의 한국프로야구 기록대백과와 Korea Baseball Organization (2010, 2011, 2012, 2013, 2014)의 한국프로야구 연감에서 구한 것이다. 본 논문의 통계분석과 그래프 작성은 모두 Minitab Release 14.12.1을 이용하여 수행되었다.

3.1. 기존의 OPS와 수정된 OPS의 연도별 비교

1982년부터 2013년까지 각 연도별로 팀들의 두 가지 타격지표(OPS와 수정OPS)와 팀의 경기당 평균득점 간의 상관계수를 계산한 결과가 Table 3.1에 나와 있다. 이 표에서 AOPS는 수정OPS를 나타낸다. 여기서 볼 수 있듯이 32개의 시즌 중 1982년, 1994년, 1996년, 1997년, 2003년, 2004년, 2009년을 제외한 25개의 시즌에 수정OPS와 경기당 평균득점 간의 상관계수가 OPS와 경기당 평균득점 간의 상관계수보다 높으며, 이것은 매우 유의한 결과이다(이항검정법의 p 값 = 0.0011). 이로부터 식 (2.2)와 같이 수정된 수정OPS를 사용하면 기존의 OPS보다 팀의 득점력을 더욱 잘 설명한다는 것을 알 수 있다.

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

	BA	OBP	Slg	IsoP	OPS	AOBP	AOPS
Corr. coeff.	0.85019	0.87818	0.91700	0.81514	0.95061	0.87917	0.95152
Rank Corr. Coeff.	0.84209	0.88260	0.91060	0.80914	0.94713	0.88314	0.94822

3.2. 몇 가지 지표와 득점력 사이의 관계

이 절부터는 1982년부터 2013년까지 연 244개 팀의 1년 단위 기록들을 근거로 하여 분석한다. 먼저 여러 가지의 타격지표와 득점력 사이의 관계를 알아본다. 팀들의 타율과 출루율, 장타율 등 여러 지표와 경기당 평균득점 간의 상관계수를 구하여 Table 3.2에 나타냈다. 또한 각각의 지표를 기준으로 하여 팀들의 순위를 구한 뒤 경기당 평균득점을 기준으로 한 순위와의 순위상관계수를 구하여 표에 함께 나타냈다. 이 표에서 BA, OBP, Slg, AOBP는 각각 타율, 출루율, 장타율, 수정출루율을 나타낸다. 또한 IsoP는 'Isolated Power'의 약자로서 '절대장타율'이라고도 하며, 다음 식으로 정의된다.

$$\text{IsoP} = \text{장타율} - \text{타율}. \quad (3.1)$$

Table 3.2에서 볼 수 있듯이 상관계수와 순위상관계수는 모든 지표에서 매우 유사한 값을 갖는다. 여러 지표들을 비교해 보면, IsoP가 평균득점과 가장 낮은 상관관계를 보이고 있으며, 타율이 두 번째로 낮은 상관관계를 보인다. 따라서 득점 생산에 대한 기여도를 논할 때 IsoP와 타율은 배제하고 출루율과 장타율에 주목하기로 하며, 출루율과 장타율을 결합한 OPS, 그리고 2절에서 정의된 수정출루율과 수정OPS에 초점을 맞추기로 한다. Table 3.2를 보면, 현재 사용되고 있는 타격지표 중 OPS가 평균득점과 가장 높은 양의 상관관계(상관계수 0.95061)를 보이고 있다. 하지만 새롭게 제시된 수정OPS가 OPS보다도 좀 더 높은 양의 상관관계(상관계수 0.95152)를 보이는 것으로 나타났다.

3.3. 가중수정OPS

현재 공신력 있는 공격지표로 사용되고 있는 OPS는 출루율과 장타율을 똑같은 비중으로 반영하고 있는데, 이와 같이 출루율과 장타율의 단순합계인 OPS가 널리 쓰이는 것은 단지 계산이 간편하다는 이유 때문이다. 그러나 이렇게 두 성분에 똑같은 비중을 주는 것에 대한 합리적 근거는 전혀 없다. 따라서 출루율과 장타율 중 어느 한 쪽에 더 큰 비중을 둔 가중평균을 사용하여 팀의 득점력을 좀 더 잘 설명할 수 있다면, 단순한 OPS보다 더욱 합리적인 타격지표가 될 것이다. 이러한 의미에서 Kim (2012)은 다음의 식으로 정의되는 가중수정OPS를 제안하였다.

$$\text{가중OPS} = a \times \text{출루율} + (1 - a) \times \text{장타율}, \quad (3.2)$$

여기서 a 는 0과 1사이의 값으로서 출루율의 가중값을 나타낸다. Kim (2012)은 2007년부터 2011년까지의 한국 프로야구 정규리그 전 경기 자료를 분석하였는데, 그 결과 출루율에 57%의 비중을 두고 장타율에 43%의 비중을 두는 가중OPS가 팀의 득점력을 가장 잘 설명하는 것으로 밝혀졌다.

본 논문에서는 위의 아이디어를 2절에서 설명한 수정OPS에 적용하고자 한다. 즉 수정출루율과 장타율의 가중평균인 가중수정OPS를 다음 식으로 정의한다.

$$\text{가중수정OPS} = a \times \text{수정출루율} + (1 - a) \times \text{장타율}. \quad (3.3)$$

a 는 역시 0과 1사이의 값으로서 수정출루율의 가중값을 나타낸다. $a = 0$ 인 경우에 가중수정OPS는 장타율이 되고, $a = 1$ 인 경우에는 수정출루율이 되며, $a = 0.5$ 인 경우에는 수정OPS와 같은 개념을 갖는

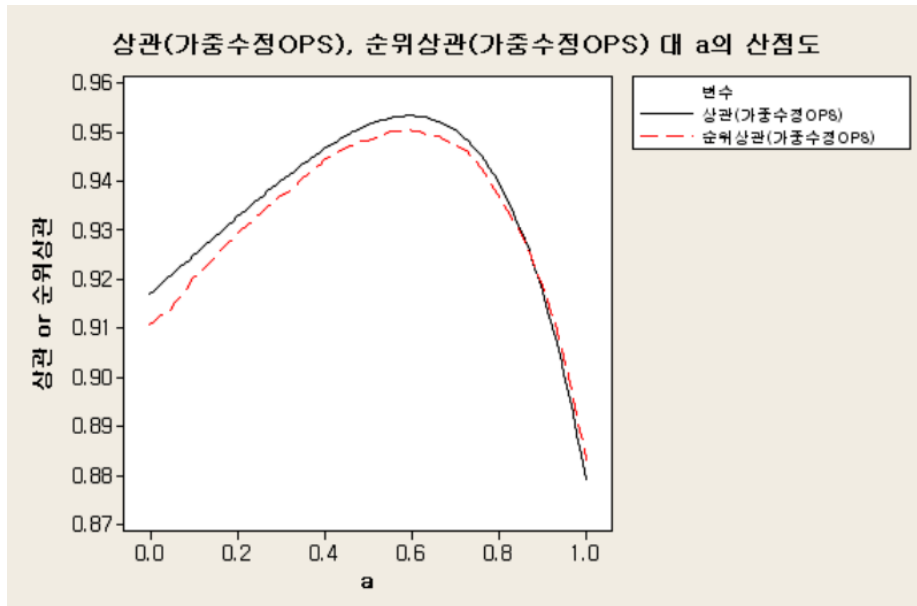


Figure 3.1. Plot of correlation coefficient and rank correlation coefficient against a (weighted AOPS)

것이 된다. a 의 값을 0부터 1까지 0.01간격으로 변화시켜 가며 팀들의 가중수정OPS를 구하고, 이 가중수정OPS와 팀의 경기당 평균득점 간의 상관계수와 순위상관계수를 구하여, a 값에 따라 변화하는 내용을 그래프로 작성한 것이 Figure 3.1과 같다. 이 그림에서 실선은 상관계수를 나타내고 점선은 순위상관계수를 나타낸다.

32개 전 시즌의 자료를 근거로 분석한 결과 상관계수는 $a = 0.60$ 일 때 최댓값을 가지며, 그 값은 0.95343인 것으로 밝혀졌다. 이것은 가중수정OPS 중에서는

$$\text{가중수정OPS} = a \times \text{수정출루율} + (1 - a) \times \text{장타율} \tag{3.4}$$

로 정의되는 지표를 사용할 때 팀들의 득점력이 가장 잘 설명된다는 것을 의미한다. 순위상관계수는 $a = 0.59$ 일 때 최댓값 0.95053을 갖는 것으로 밝혀졌고, $a = 0.60$ 일 때는 0.95044로서 두 번째로 큰 값이었다. 상관계수를 최대로 하는 값과 순위상관계수를 최대로 하는 값에 미세한 차이가 있으나, 지표의 값들로부터 직접 계산된 상관계수가 우선이므로 식 (3.4)로 정의된 지표를 사용하는 것이 타당할 것이다.

Figure 3.2는 a 의 값이 0.3에서 0.8까지의 범위에 있을 때 식 (3.2)의 가중OPS와 평균득점 간의 상관계수, 그리고 식 (3.3)의 가중수정OPS와 평균득점 간의 상관계수를 산점도로 겹쳐서 그린 것이다. 이 범위의 모든 값에서 가중수정OPS가 가중OPS보다 높은 상관관계를 보이는 것을 알 수 있다. 한편 Figure 3.3은 a 의 값이 0.3에서 0.8까지의 범위에 있을 때 식 (3.2)의 가중OPS와 평균득점 간의 순위상관계수, 그리고 식 (3.3)의 가중수정OPS와 평균득점 간의 순위상관계수를 산점도로 겹쳐서 그린 것이다. 이 범위의 모든 값에서 가중수정OPS가 가중OPS보다 높은 순위상관계수를 갖는 것을 알 수 있다. 참고로, 가중OPS는 $a = 0.60$ 일 때 최대의 순위상관계수 0.94969를 갖는 것으로 계산되었는데, 가중수정OPS의 최대순위상관계수는 앞에서 밝혔듯이 $a = 0.59$ 일 때 0.95053으로서 좀 더 높았다.

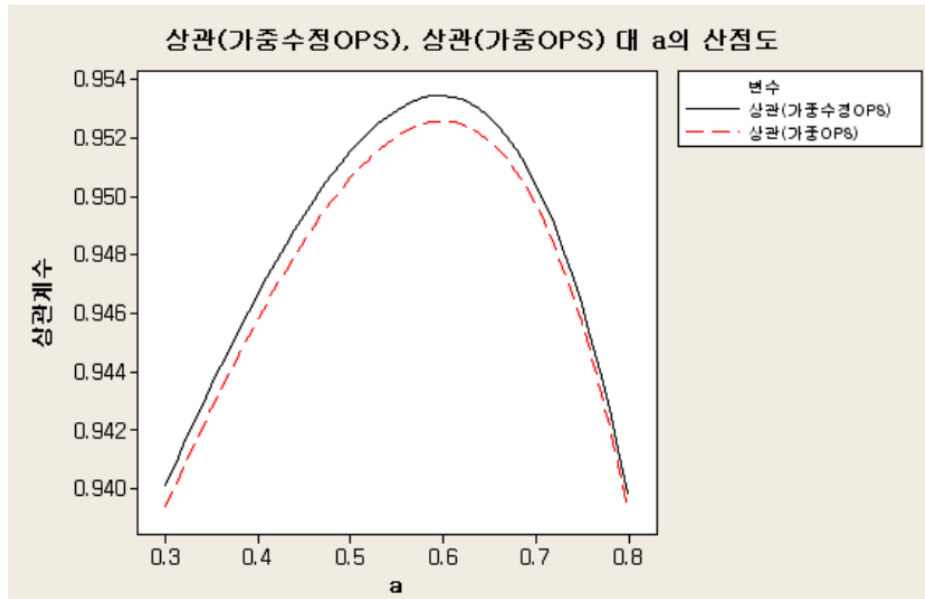


Figure 3.2. Plot of correlation coefficient of weighted AOPS and weighted OPS with average runs per game against a

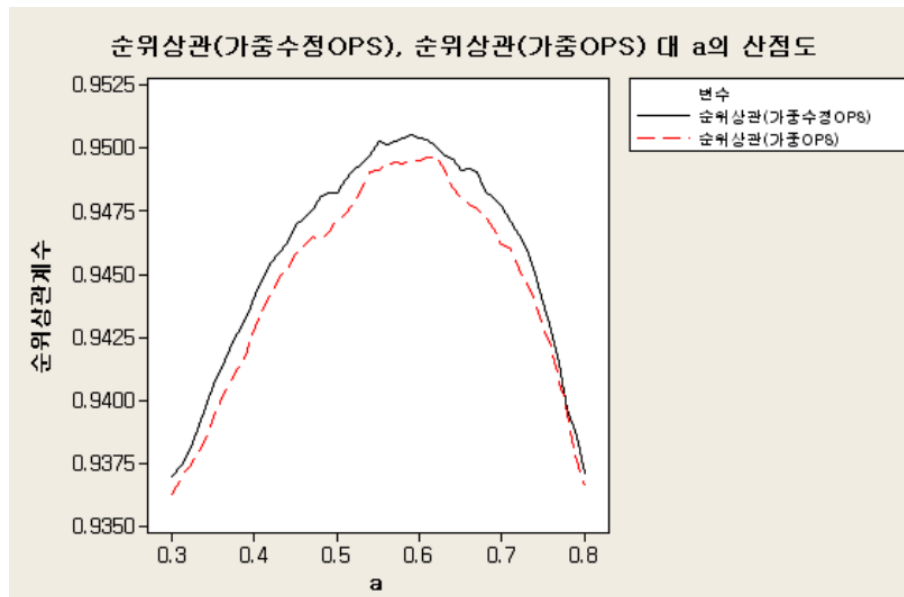


Figure 3.3. Plot of rank correlation coefficient of weighted AOPS and weighted OPS with average runs per game against a

3.4. 논의

3.1절부터 3.3절까지의 내용을 종합하면, 팀의 득점력을 가장 잘 설명해 주는 타격지표는 식 (3.4)로 표시된 가중수정OPS(0.60)인 것으로 나타났으며, 이 지표를 사용할 때 팀의 경기당 평균득점과의 상관

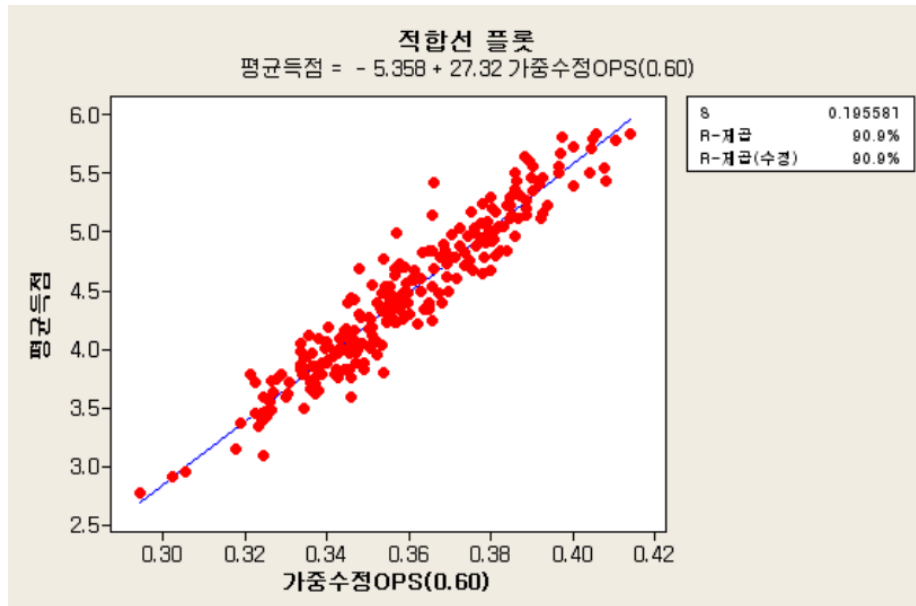


Figure 3.4. Plot of average run against weighted AOPS with $a = 0.60$ (1982 ~ 2013)

계수는 0.95343, 순위상관계수는 0.95044로 매우 높았다. 참고로, 수정되지 않은 기존의 출루율을 바탕으로 한 가중OPS를 사용할 때의 상관계수를 구해 보니 $a = 0.60$ 일 때 최댓값 0.95257을 갖는 것으로 나타났으며, 이때 순위상관계수는 0.94954였다. 즉 기존의 출루율을 바탕으로 한 가중OPS를 사용할 때보다 수정출루율을 바탕으로 한 가중수정OPS를 사용할 때 득점에 대한 설명력이 더욱 높아진다는 것을 알 수 있다. 수정출루율의 상관계수(0.87917)보다 장타율의 상관계수(0.91700)가 더 높지만 최적의 가중수정OPS에서는 수정출루율에 더 큰 가중값(60%)이 붙는다는 점이 주목할 만하다. 가중수정OPS(0.60)과 팀의 경기당 평균득점 간의 산점도를 최소제곱회귀직선과 함께 나타낸 것이 Figure 3.4이다. 가중수정OPS(0.60)을 x , 평균득점을 y 라 할 때, 최소제곱회귀직선의 방정식은

$$\hat{y} = -5.358 + 27.317x \tag{3.5}$$

였다. 결정계수는 90.9%로 매우 높았으며, 회귀분석의 F 통계량(자유도 1과 242)의 값은 2418.46이었고 p 값은 논하는 것이 무의미할 정도로 작았다.

Table 3.3은 식 (3.5)에 근거를 두고 표준화잔차의 절댓값이 큰 비정상적인 관측값들을 찾아낸 것이다(이 표에서 RPG는 경기당 평균득점을 나타낸다). 실제 득점이 적합치보다 특별히 많이 나온 팀들은 1982년의 삼성, 1995년의 OB, 1995년의 롯데, 2008년의 두산, 2005년의 LG였으며, 그 반대의 경우는 1987년의 빙그레, 1988년의 태평양, 1986년의 빙그레, 2003년의 두산이었다.

이번에는 두 개의 회귀모형(모형 1: 출루율과 장타율을 설명변수로 하고 경기당 평균득점(y)을 반응변수로 하는 중회귀모형, 모형 2: 수정출루율과 장타율을 설명변수로 하고 경기당 평균득점(y)을 반응변수로 하는 중회귀모형)을 적합해 보았다. 그 결과는 다음과 같이 모형 2가 좀 더 좋은 결과를 주는 것으로 나타났다.

모형 1: $\hat{y} = -5.381 + 16.472 \times \text{출루율} + 11.022 \times \text{장타율}$ ($\sqrt{\text{MSE}} = 0.198$, 결정계수 = 0.907).

Table 3.3. Unusual observations based on equation (3.5)

Year	Team	AOPS(0.60)	RPG	\widehat{RPG}	$\widehat{SE}(\widehat{RPG})$	standardized residual
1982	Samsung	0.366	5.4250	4.6429	0.0132	4.01
1995	OB	0.357	4.9921	4.3900	0.0126	3.08
1995	Lotte	0.348	4.6825	4.1383	0.0140	2.79
2008	Doosan	0.366	5.1349	4.6261	0.0131	2.61
2005	LG	0.354	4.7698	4.3084	0.0128	2.36
1987	Binggrae	0.354	3.7963	4.3048	0.0128	-2.61
1988	Pacific	0.346	3.5833	4.0849	0.0145	-2.57
1986	Binggrae	0.324	3.0833	3.5013	0.0228	-2.15
2003	Doosan	0.366	4.2331	4.6292	0.0131	-2.03

모형 2: $\hat{y} = -5.339 + 16.270 \times \text{수정출루율} + 10.982 \times \text{장타율} (\sqrt{\text{MSE}} = 0.196, \text{결정계수} = 0.909)$.

가중수정OPS(0.60)이 팀의 득점력과 가장 밀접한 관련이 있으므로 타자들의 공격 기여도도 이 지표를 기준으로 평가하는 것이 기존의 OPS를 기준으로 평가하는 것보다 합리적이라고 사료된다. Table 3.4는 최근의 4개 시즌인 2010년부터 2013년까지 연도별로 가중수정OPS(0.60)을 지표로 하여 규정타석을 채운 타자들의 순위를 1위부터 10위까지 나타낸 것이다. 표에서 W-AOPS(0.60)이 가중수정OPS(0.60)을 나타낸다. 이 순위를 팀의 득점에 기여한 정도로 해석할 수도 있을 것이다. 함께 표시된 OPS-R은 기존의 OPS를 기준으로 한 순위이며, W-OPS(0.60)-R은 Kim (2012)에서 제시된 가중OPS에 근거를 둔 가중OPS(0.60)을 기준으로 한 순위이다. 세 지표에 의한 결과에 곳곳에서 차이가 있는 것을 볼 수 있다. 특히 2010년과 2013년에는 5위권 이내에서도 가중수정OPS(0.60)과 기존의 OPS 간에 순위의 변동이 있었으며, 2010년에는 가중수정OPS(0.60)과 가중OPS(0.60) 간에도 순위의 변동이 생겼다.

4. 결론

다양한 야구 데이터의 분석 결과는 팀의 득점력과 깊은 관련이 있는 타격지표가 출루율과 장타율임을 말해 준다. 타격지표 중 가장 간단하고 기본적인 것은 타율이지만, 이것은 4사구를 고려하지 않은 것이므로 득점 설명 능력이 출루율보다 떨어진다. 뿐만 아니라 타율은 루타수를 고려하지 않은 것이므로 득점 설명 능력이 장타율보다도 떨어진다. 3.2절의 결과도 이를 잘 보여 주고 있다. 이런 이유로 출루율과 장타율, 그리고 이들을 결합한 OPS가 팀과 개인의 타격 능력을 나타내는 지표로 중요하게 사용되고 있다.

본 논문에서는, 타율과 출루율에서 희생플라이를 처리하는 방식이 다르다는 점에 주목하여, 현재 사용되고 있는 출루율을 정의하는 식의 분모에서 희생플라이수를 제거함으로써 수정출루율을 정의하고 이를 바탕으로 수정OPS를 정의하였다. 이 수정OPS는 역대 한국 프로야구의 32개 시즌 중 25개의 시즌에서 기존의 OPS보다 팀들의 득점력을 더 잘 설명하였다. 수정OPS는 32개 시즌에 걸쳐 연 244개 팀의 1년 단위 기록들을 근거로 하여 분석한 결과에서도 기존의 OPS보다 더 높은 득점 설명력을 보였다. 본 논문에서는 더 나아가서 수정출루율과 장타율에 여러 가지의 가중값을 주는 가중수정OPS를 정의하였으며, 원년인 1982년부터 2013년까지의 한국 프로야구 정규리그 전 경기 자료를 바탕으로 팀의 득점력을 가장 잘 설명하는 최적의 가중수정OPS를 구하였다. 구체적으로 수정출루율에 60%, 장타율에 40%의 비중을 두는 가중수정OPS가 팀의 득점을 가장 잘 설명하는 것으로 나타났으며, 경기당 평균득점과의 상관계수는 0.95343이었다. 이 가중수정OPS는 Kim (2012)에서 제안된 가중OPS보다 득점력을 더욱

Table 3.4. Top 10 batters based on weighted AOPS(0.60) for each year(2010 ~ 2013)

2010					
Rank	Batter	Team	W-AOPS(0.60)	OPS-R	W-OPS(0.60)-R
1	LeeDH	Lotte	0.535	1	1
2	HongSH	Lotte	0.499	2	2
3	KimHS	Doosan	0.464	3	3
4	ParkSM	Samsung	0.459	7	4
5	ChoiJS	Doosan	0.456	4	5
6	ChoiJ	SK	0.455	6	6
7	KimDJ	Doosan	0.4496	8	8
8	ChoIS	LG	0.4495	5	7
9	ChoiHW	Samsung	0.445	9	9
10	ParkJK	SK	0.438	11	10
2011					
Rank	Batter	Team	W-AOPS(0.60)	OPS-R	W-OPS(0.60)-R
1	ChoiHW	Samsung	0.508	1	1
2	LeeDH	Lotte	0.494	2	2
3	ChoiJ	SK	0.455	3	3
4	SonAS	Lotte	0.436	4	4
5	KimDJ	Doosan	0.428	5	5
6	LeeBK	LG	0.421	6	6
7	LeeYK	KIA	0.419	9	7
8	KimHS	Doosan	0.417	8	8
9	KangMH	Lotte	0.412	7	9
10	ChoiJH	Hanwha	0.404	10	10
2012					
Rank	Batter	Team	W-AOPS(0.60)	OPS-R	W-OPS(0.60)-R
1	KimTK	Hanwha	0.501	1	1
2	KangJH	Nexen	0.474	2	2
3	ParkSM	Samsung	0.472	3	3
4	ParkBH	Nexen	0.463	4	4
5	ChoiJ	SK	0.4490	5	5
6	JeongSH	LG	0.4486	6	6
7	LeeHJ	SK	0.440	7	7
8	LeeSY	Samsung	0.433	8	8
9	KimWS	KIA	0.407	12	9
10	HongSH	Lotte	0.406	9	10
2013					
Rank	Batter	Team	W-AOPS(0.60)	OPS-R	W-OPS(0.60)-R
1	ParkBH	Nexen	0.506	1	1
2	ChoiJ	SK	0.479	2	2
3	ParkSM	Samsung	0.462	3	3
4	KimTK	Hanwha	0.459	4	4
5	SonAS	Lotte	0.443	7	5
6	ParkJK	SK	0.440	6	6
7	ChoiHW	Samsung	0.445	5	7
8	KangJH	Nexen	0.430	8	8
9	NaJW	KIA	0.4283	9	9
10	MinBH	Doosan	0.4280	10	10

잘 설명해 주었다.

본 논문에서는 우리나라의 프로야구 자료를 근거로 연구하였다. 팀별로 성향이 다를 수도 있으므로 팀에 따른 연구도 생각할 수 있을 것이다. 또한 미국이나 일본의 프로야구에 관심을 가지고 있는 연구자가 연구한다면, 각국의 야구 성향과 나라 간의 차이를 알아볼 수도 있는 작업이 될 것이다.

References

- 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.
- Kim, H. J. (2012). Effects of on-base and slugging ability on run productivity in Korean professional baseball, *Journal of the Korean Data & Information Science Society*, **23**, 1065–1074.
- Korea Baseball Organization (2009). *The Official Baseball Encyclopedia 2009*, Korea Baseball Organization, Seoul.
- Korea Baseball Organization (2010). *2010 Official Baseball Guide*, Korea Baseball Organization, Seoul.
- Korea Baseball Organization (2011). *2011 Official Baseball Guide*, Korea Baseball Organization, Seoul.
- Korea Baseball Organization (2012). *2012 Official Baseball Guide*, Korea Baseball Organization, Seoul.
- Korea Baseball Organization (2013). *2013 Official Baseball Guide*, Korea Baseball Organization, Seoul.
- Korea Baseball Organization (2014). *2014 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. T. (2014a). Measurements for hitting ability in the Korean pro-baseball, *Journal of the Korean Data & Information Science Society*, **25**, 349–356.
- Lee, J. T. (2014b). Estimation of OBP coefficient in Korean professional baseball, *Journal of the Korean Data & Information Science Society*, **25**, 357–363.
- 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.

한국 프로야구에서 가중수정OPS를 이용한 득점력 설명

김혁주^{a,1} · 김예형^a

^a원광대학교 수학·정보통계학부

(2014년 7월 21일 접수, 2014년 9월 18일 수정, 2014년 10월 2일 채택)

요약

한국 프로야구에서 장타율과 수정된 출루율을 이용하여 팀의 득점력을 설명하기 위한 지표인 수정OPS와 가중수정OPS를 제시하였다. 먼저 현재 사용되고 있는 출루율을 수정한 수정출루율을 정의하였으며, 수정OPS를 수정출루율과 장타율의 합으로, 가중수정OPS를 수정출루율과 장타율의 가중평균으로 정의하였다. 한국 프로야구 원년인 1982년부터 2013년까지의 정규리그 전 경기 자료를 분석한 결과 수정OPS는 기존의 OPS보다 득점력을 더 잘 설명해 줬다. 연도별 비교에서도 전체 32개 시즌 중 25개의 시즌에서 수정OPS가 기존의 OPS보다 득점력을 더 잘 설명했다. 32개 시즌의 자료를 종합한 결과 수정출루율에 60%, 장타율에 40%의 가중값을 주는 가중수정OPS가 팀의 득점력을 가장 잘 설명하는 타격지표인 것으로 밝혀졌다. 이 가중수정OPS는 Kim (2012)에서 제시된 가중OPS보다도 더욱 높은 설명력을 갖는 것으로 밝혀졌다.

주요용어: 수정출루율, 수정OPS, 가중수정OPS, 득점, 상관계수.

이 논문은 2012학년도 원광대학교의 교비 지원에 의해서 수행됨.

¹교신저자: (570-749) 전북 익산시 익산대로 460, 원광대학교 수학·정보통계학부 및 기초자연과학연구소.

E-mail : hjkim@wonkwang.ac.kr