

한국프로야구에서 출루율 계수의 추정

이장택¹

¹단국대학교 응용통계학과

접수 2014년 1월 18일, 수정 2014년 2월 8일, 게재확정 2014년 2월 26일

요약

타자들의 타격능력을 나타내는 OPS를 계산할 때 출루율을 단순하게 장타율과 더한다. 하지만 실제로 출루율에 가중치를 부여하여 계산한 OPS가 게임당 평균득점과 상관관계가 더 커지게 되는데, 본 연구에서는 한국프로야구에 있어서 가장 적합한 가중치를 전체 데이터 및 연대별 데이터를 이용하여 추정하였다. 제안된 가중치는 게임당 평균득점의 영향을 받으며, 가중치와 게임당 득점과의 관계는 회귀직선으로 설명하였다.

주요용어: 상관계수, 세이버메트릭션, 오피에스, 장타율, 출루율, 출루율 계수.

1. 머리말

야구데이터 분석가인 세이버메트릭션 (sabermetrician)들이 고안해 놓은 각종 지표 중에서 타자들을 평가하는 통계량으로 가장 많이 사용되는 것은 아마도 OPS (on-base plus slugging)일 것이다. OPS는 야구에서 타자들을 평가하는 통계량으로 타율만으로 제대로 평가하기 어려운 장타자들을 평가할 때 유용하게 사용되는데, 2000년대 후반부터는 우리나라 야구 중계에서도 자주 등장한다. OPS는 출루율 OBP (on base percentage)과 장타율 SLG (slugging percentage)의 합으로 계산된다. OPS는 2010년부터는 KBO 홈페이지에도 게재되기 시작했으며 타자의 출루능력과 장타력을 골고루 평가하며 누구나 쉽게 이해하고 계산할 수 있다는 장점이 있다. OBP와 SLG는 각각 다음과 같이 정의되는데, 이 경우 H는 안타, BB는 볼넷, HBP는 사구, AB는 타수, SF는 희생타, 1B는 1루타, 2B는 2루타, 3B는 3루타, HR은 홈런을 각각 의미한다.

$$OBP = (H + BB + HBP)/(AB + BB + HBP + SF) \quad (1.1)$$

$$SLG = [1B + 2(2B) + 3(3B) + 4(HR)]/AB \quad (1.2)$$

한국프로야구의 경우, OPS에 관한 직접적인 연구로 Cho와 Cho (2005)는 득점과 OPS의 관계를 분석하고, OPS를 구성하는 OBP와 SLG를 이용하여 군집분석을 실시하고 각 군집의 경기당 득점을 비교분석하였다. 또한 Kim (2012)은 2007년부터 2011년까지의 프로야구 데이터를 이용하여 여러 가지 가중값을 갖는 타격지표들이 팀의 경기당 평균득점과 갖는 상관관계를 분석하였으며 그 결과 출루율에 57%, 장타율에 43%의 가중값을 주는 가중OPS, 즉 출루율 계수가 1.33인 경우에 득점 생산성을 가장 잘 설명하는 타격지표인 것으로 밝혔다. 또한 한국프로야구에 대한 최근 연구로는 시계열모형을 이용하여 관중 수 예측을 다룬 Lee와 Bang (2010), 한국프로야구 타자들에 대한 세이버메트릭스 지수 값을 이용하여 선수들의 경기력과 연봉간의 패턴을 분석한 Seung과 Kang (2012) 등이 있다. 한편 해외연구로

¹ (448-701) 경기도 용인시 죽전동 126번지, 단국대학교 응용통계학과, 교수. E-mail: jtlee@dankook.ac.kr

는 GPA를 제안한 Aaron (2003)과 1960년 이후의 메이저리그 데이터를 이용하여 출루율 계수를 구한 Wang (2006)이 있다. OPS는 단순히 OBP와 SLG를 더해서 만들어지는 식이기 때문에 계산 및 기억하는 데는 편리하지만 2개의 통계량을 동일한 가치로 취급하는 것은 사실 문제가 된다. 왜냐하면 실제로 OBP와 SLG의 평균과 표준편차의 차이가 엄연히 존재하며 또 분포의 차이도 있다. 이와 같은 이유로 세이버메트릭스들은 타자의 최종목표가 팀이 득점하는 것이므로 OBP와 SLG의 가중치 부여에 대해 OPS와 득점과의 상관관계를 통해 연구하기 시작했다. 그 결과, 인터넷에서 살펴볼 수 있는 모든 연구에서 OPS를 계산할 때 OBP의 가중치가 SLG의 가중치보다 커야 한다는 결론을 내렸다. 이와 같은 철학을 이용하여 만들어진 통계량이 GPA (gross production average)인데, OBP에 1.8의 가중치를 부여한 것이다. 여기서 4로 나누는 이유는 타율과 비슷한 값이 나오도록 하기 위한 숫자로 GPA의 값이 3할이 나오면 뛰어난 타자라고 할 수 있겠다.

$$GPA = (1.8 \times OBP + SLG)/4 \quad (1.3)$$

식 (1.3)에서 계수 1.8은 소수점 1자리 숫자 중 1960년 이후의 메이저리그 데이터를 이용하여 구한 게임당 평균득점 (runs per game; RPG)과 가장 상관관계를 크게 만드는 OBP의 가중치다 (Aaron, 2003; Wang, 2006). 따라서 숫자를 소수점을 늘리거나 사용하는 데이터를 바꾸면 GPA계수는 다른 숫자의 값으로 바뀔 수도 있다. 이와 같은 이유로 본 연구에서는 한국프로야구에서 가장 타당한 OBP에 곱해지는 가중치인 출루율 계수의 값을 게임당 득점과의 상관관계를 통해서 알아보고, 최적 출루율 계수의 값을 전체 및 연대별로 구해본다. 또한 최적 출루율 계수와 게임당 득점과의 관계를 회귀분석을 통해 최적 출루율 계수의 값이 변하는 주된 이유를 탐색하여 본다.

본 논문은 다음과 같이 구성되어 있다. 2절에서는 한국프로야구에 있어서 OPS, OBP, SLG에 관한 기술통계와 가중치를 부여한 OPS를 설명하였으며, 3절에서는 연대별 및 전체 데이터를 이용하여 최적 출루율 계수의 값을 추정하고 한국프로야구에서의 최적 출루율 계수의 변화를 살펴봄으로써 최적 출루율 계수를 사용하여 구한 wOPS의 역대 팀 및 개별 선수들의 상위 10위 이내기록을 요약하였으며, 게임당 평균득점 (RPG)과 최적 출루율 계수 사이의 회귀직선을 제시하였다. 끝으로 4절에서는 본 연구의 결론에 대해 언급하였다.

2. 연구방법

2.1. 데이터의 구성

본 연구에 사용된 데이터는 한국야구위원회 (KBO)에 기록되어 있는 1982년부터 2013년까지의 각 팀의 정규리그 전 경기결과이다. 통계패키지 SPSS 18K와 수집된 데이터를 이용하여 기술통계량을 계산하였다. 고려된 데이터는 모두 244개이며 팀의 이름이 한 번도 바뀌지 않은 삼성과 롯데가 32번 등장하여 최다빈도수를 나타냈다. 또한 타자들에 대한 개인기록에 대해서는 2000년부터 2013년 사이에 있었던 경기 중 규정타석 수를 채운 타자 594명의 경기결과를 사용하였다.

Table 2.1 Mean of OBP, SLG and OPS (standard deviations in parenthesis)

Year	OBP	SLG	OPS
1982-1989	0.332 (0.014)	0.368 (0.029)	0.700 (0.041)
1990-1999	0.333 (0.017)	0.383 (0.037)	0.717 (0.052)
2000-2013	0.344 (0.014)	0.396 (0.032)	0.741 (0.043)
1982-2013	0.338 (0.016)	0.386 (0.035)	0.724 (0.049)

Table 2.1은 타자 통계량 OBP, SLG, OPS의 평균값 및 표준편차를 전체 및 연대별로 나타낸 결과이다. 전체 및 연대별의 모든 경우에 SLG는 OBP보다 평균이 크며 표준편차도 크다. 또한 한국프로야구의 평균 OPS는 0.7 초반에서 형성되는 사실을 확인할 수 있다.

2.2. 가중치를 부여한 OPS

득점 생산성을 잘 나타내게 하기 위하여 가중치인 출루율 계수 (OBP coefficient) ω 를 곱한 OPS (wOPS)를 식 (2.1)과 같이 정의한다. 여기서 $\omega = 1$ 이면 기존의 OPS가 된다.

$$wOPS = \omega \times OBP + SLG, \quad 0 < \omega < \infty. \tag{2.1}$$

일반적으로 야구에서는 상대방보다 더 많은 득점을 해야 이기는 것이므로 타자의 능력을 측정하는 가장 좋은 방법은 팀의 득점에 얼마만큼 공헌을 했느냐로 보통 설명한다. 그래서 출루율 계수 ω 를 결정하는 방법은 보통 게임당 득점 (RPG)이 전체 득점보다 wOPS와 상관관계가 크기 때문에 RPG와 wOPS의 상관관계수가 가장 큰 경우의 ω 를 사용한다. 이 경우 득점을 R , 참가경기 수를 G 로 표기할 때 $RPG = R/G$ 로 정의된다.

3. 분석결과 및 논의

3.1. 최적 출루율 계수의 값

OPS는 가중치의 부여 여부와 상관없이 득점 (R)보다는 게임당 득점 (RPG)과 상관관계가 크다. 실제로 전체 프로야구 데이터를 이용하여 계산하면 R과 OPS의 상관관계수는 0.858, RPG와 OPS의 상관관계수는 0.950으로 상당한 차이가 난다. 이런 이유로 본 연구에서는 ω 의 값을 0.01부터 3.00까지 0.01 간격으로 변화시켜가며 RPG와 wOPS의 상관관계수를 계산하였다. 그 결과 1982년부터 2013년까지 한국 프로야구 전체 데이터를 이용하여 RPG와 wOPS의 상관관계수를 구하면 ω 가 1.47인 경우에 최댓값을 갖는다는 사실을 알 수 있다. 이 값은 Kim (2012)의 2007년부터 2011년까지의 정규리그 전 경기 타격 자료를 이용하여 구한 1.33보다 큰 값이다. 하지만 ω 를 1.47로 사용하여 구한 wOPS는 연대별로 RPG와의 상관관계수를 구하면 1980년대는 0.914, 1990년대는 0.966, 2000년 이후에는 0.948로 1980년대가 현저하게 상관관계가 약한데, 이는 1980년대는 게임 수도 적었고 프로야구 초창기라 득점이 OPS 이외의 요인인 실책 등으로 인한 경우도 많았기 때문이라고 생각할 수 있다. 따라서 1980년대 데이터를 생략하고 1990년대 이후 데이터로 RPG와 wOPS의 상관관계수를 구하면 ω 가 1.56인 경우에 최댓값을 갖는다는 사실을 알 수 있고, 이것은 메이저리그의 1.8에 좀 더 가까운 수치이다. 전체 데이터와 연대별 데이터를 이용하여 구한 최적 출루율 계수 ω 의 값은 Table 3.1과 같다. 따라서 출루율 계수 ω 에 관한 추론이 본 연구의 목적이므로 이후부터는 ω 의 추정값으로 1.56을 사용한다.

Table 3.1 Value of optimal weight for four year cases

Year	Optimal weight ω
1982-1989	1.23
1990-1999	1.72
2000-2013	1.70
1982-2013	1.47
1990-2013	1.56

wOPS ($\omega = 1.56$)을 x , RPG를 y 라고 할 때, 90년대 이후의 데이터 192개를 이용하여 추정된 회귀직선

의 방정식 \hat{y} 은 다음과 같으며,

$$\hat{y} = -5.296 + 10.677x \quad (3.1)$$

결정계수는 91.6%로 높았으며, 분산분석의 p 값은 $p < 0.001$ 로 회귀직선은 유의수준 1%에서 매우 유의한 것으로 나타났다. 또한 더빈-왓슨 통계량의 값은 1.584로 검정의 임계값과 유사하여 오차는 자기상관의 영향이 거의 없음을 알 수 있었는데, 이 경우 사용된 데이터가 시간의 흐름에 따라 일정한 간격으로 기록되어졌지만 타자들의 능력이 시간에 따라 상승 및 하락을 한다고는 판단되어지지 않기에 본 연구에서 시간의 흐름을 배제하였으나, wOPS의 변화를 시계열분석을 이용하여 연구를 하는 것도 의미가 있을 것으로 간주된다.

3.2. 최적 출루율 계수와 RPG의 관계

출루율 계수 ω 는 RPG와 wOPS의 상관관계를 최대화하는 양수이므로 RPG의 평균과 연관이 있다. 따라서 ω 가 크면 RPG 역시 클 가능성이 많다. 이 사실을 좀 더 확실하게 입증하기 위하여 wOPS와 RPG의 연관성이 약한 1980년대 데이터를 배제하고 1990년 이후만 사용하여 2013년까지의 데이터 192개를 이용하여 RPG의 값을 오름차순으로 정리하였다. 그리고 하위 40개, 하위 45개, 하위 50개 등 5개를 추가하여 29개의 데이터셋을 생성한 후 각각의 데이터셋을 이용하여 RPG의 평균과 최적 출루율 계수 ω 의 값을 계산하였다.

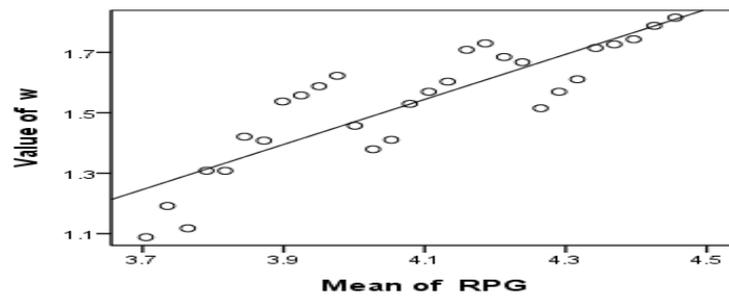


Figure 3.1 Plot of optimal OBP coefficient against mean of RPG

Figure 3.1은 RPG의 평균과 최적 출루율 계수 ω 의 값 간의 산점도와 최소제곱회귀직선을 함께 나타낸 것이다. RPG의 평균을 m 이라 표기하면 최소제곱회귀직선의 방정식은 다음과 같이 구하여진다.

$$\hat{\omega} = -1.514 + 0.746m \quad (3.2)$$

이 경우 분산분석의 p 값은 $p < 0.001$ 로 회귀직선은 유의수준 1%에서 매우 유의한 것으로 나타났으며, 결정계수도 74.7%로 높았다. 또한 m 과 ω 의 상관계수는 0.864로 나타났다.

3.3. 최적 출루율 계수의 변화

최적 출루율 계수의 연도별 변화를 살펴보는 것은 해당연도 표본개수의 절대부족으로 신뢰도가 크게 떨어지기 때문에 대신 최근 5개년도 관측값들만 이용하여 최적 출루율 계수를 구하였다. Figure 3.2는 최근 5년간의 데이터만을 고려하여 구한 최적 출루율 계수와 RPG의 평균을 보여준다. 최적 출루율 계수값에 대한 주기가 5년인 단순이동평균선은 1994년까지 계속 증가하다가 1996년까지 일시적인 하락세에 접어든다. 그 뒤 2002년까지 계속 상승세를 타다가 2002년의 최고점을 뒤로 하고 현재까지 계속 하

락 추세에 있다고 할 수 있다. RPG의 평균값과 최적 출루율 계수값에 대한 주기가 5년인 단순이동평균선은 서로 비슷한 모양을 가지며, RPG의 평균값과 출루율 계수가 서로 연관이 있음을 확인시켜준다. 한편 최적 출루율 계수는 RPG보다 변화율이 크게 나타나며, 최적 출루율 계수와 RPG 평균의 주기가 5년인 단순이동평균값에 대한 상관계수는 0.769, p 값은 $p < 0.001$ 로 유의수준 1%에서 매우 유의한 것으로 나타났다.

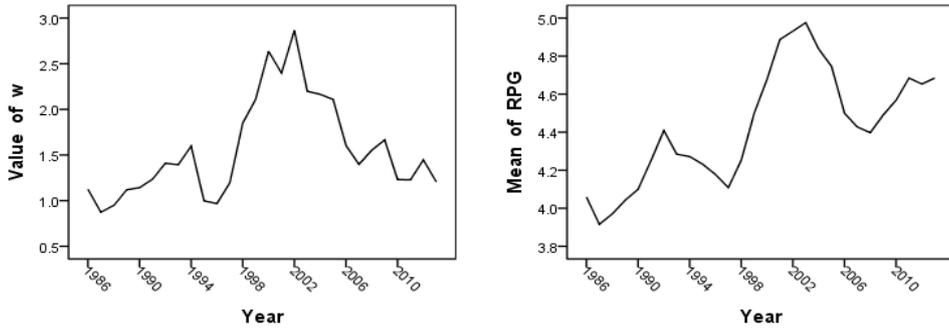


Figure 3.2 Graph for OBP coefficient and mean of RPG with 5-year observation

3.4. wOPS에 의한 역대 10위 기록

Table 3.2는 1990년부터 2013년까지 연도별로 wOPS ($\omega = 1.56$)를 이용하여 구한 팀들의 순위를 1위부터 10위까지 나타낸 것이다. 역대 10위 팀 기록은 대부분 2000년 근방에서 작성된 것을 확인할 수 있으며, 한 시즌 최고의 wOPS는 2000년도의 현대 유니콘스였다. 팀별 wOPS의 순위는 괄호 안 숫자인 OPS의 순위와 차이가 있음을 확인할 수 있다. Table 3.3은 2000년 이후 한국프로야구에서의 개인별 역대 wOPS 상위 10명을 보여주는데, 최고의 wOPS는 2001년 호세의 1.480이며, 팀과 마찬가지로 wOPS의 순위는 기존 OPS의 순위와 차이가 있음을 확인할 수 있다.

Table 3.2 Team listing of the top 10 wOPS ($\omega = 1.56$) scores

Rank	Name	Year	OBP	SLG	OPS	wOPS
01	Hyundai	2000	0.365	0.483	0.848 (01)	1.052
02	Samsung	2003	0.360	0.482	0.842 (02)	1.043
03	Hyundai	2003	0.373	0.456	0.829 (05)	1.037
04	Hanwha	1999	0.353	0.484	0.837 (03)	1.034
05	Samsung	2002	0.360	0.472	0.832 (04)	1.033
06	Haitai	1999	0.362	0.466	0.828 (06)	1.030
07	SK	2009	0.369	0.453	0.822 (08)	1.028
08	Lotte	1999	0.363	0.461	0.824 (07)	1.027
09	Samsung	1997	0.356	0.462	0.818 (09)	1.017
10	Lotte	2010	0.352	0.461	0.813 (10)	1.010

Table 3.3 Player listing of the top 10 wOPS ($\omega = 1.56$) scores

Rank	Name	Year	OBP	SLG	OPS	wOPS
01	Jose	2001	0.503	0.695	1.198 (01)	1.480
02	Sim, jeongsu	2003	0.478	0.720	1.197 (02)	1.465
03	Lee, seungyeop	2002	0.436	0.689	1.125 (04)	1.369
04	Lee, seungyeop	2003	0.428	0.699	1.127 (03)	1.367
05	Lee, daeho	2010	0.444	0.667	1.111 (05)	1.360
06	Brumbaugh	2004	0.468	0.608	1.076 (06)	1.339
07	Sim, jaehak	2001	0.473	0.599	1.072 (07)	1.337
08	Lee, daeho	2007	0.453	0.600	1.053 (08)	1.307
09	Petagine	2009	0.468	0.575	1.043 (10)	1.305
10	Kim, Dongjoo	2009	0.455	0.593	1.049 (09)	1.304

4. 결론

OPS는 간단하게 구할 수 있으면서도 타자의 득점 기여도를 잘 나타내는 지수이기 때문에 최근 들어 타자를 평가하는데 중요한 기준이 되는 통계량이다. 하지만 OBP보다 SLG의 값이 일반적으로 크므로 단순히 두 값을 더한다면 당연히 SLG 값이 큰 선수가 유리하다. 그런데, OPS가 출루를 잘 하면서 더 많은 진루를 할 수 있는 선수를 평가하기 위함이 그 목적이라면 출루율에 가중값을 부여하는 것이 타당하다. 본 연구에서는 한국프로야구 전체 데이터 중 신뢰도를 높이기 위하여 1990년 이후의 데이터를 이용하여 최적 출루율 계수를 계산하였는데, 1.56으로 나타났다. 이는 미국 메이저리그의 1.8보다 0.24 정도 작은 값인데, 주된 이유는 메이저리그보다 경기당 팀의 평균 득점이 작기 때문이라고 할 수 있겠다.

References

- Aaron, G. (2003). http://aarongleeman.com/2003_11_23_baseballblog_archive.html.
- Cho, Y. S. and Cho, Y. J. (2005). A study on OPS and runs from Korean baseball league. *Journal of the Korean Data Analysis Society*, **7**, 221-231.
- 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 (2001-2006). *2000-2005 official baseball guide*, Korea Baseball Organization, Seoul.
- Korea Baseball Organization (2014). <http://www.koreabaseball.com/Record>.
- Lee, J. T. and Bang, S. Y. (2010). Forecasting attendance in the Korean professional baseball league using GARCH models. *Journal of the Korean Data & Information Science Society*, **21**, 1041-1049.
- Seung, H. B. and Kang, K. H. (2012). A study on relationship between the performance of professional baseball players and annual salary. *Journal of the Korean Data & Information Science Society*, **23**, 285-298.
- Wang, V. (2006). The OBP/SLG ratio: What does history say? *By the Numbers*, August, 3-4.

Estimation of OBP coefficient in Korean professional baseball

Jang Taek Lee¹

¹Department of Applied Statistics, Dankook University

Received 18 January 2014, revised 8 February 2014, accepted 26 February 2014

Abstract

OPS is a sabermetric baseball statistic calculated as the sum of a player's on base percentage (OBP) and slugging percentage (SLG). One of the frequently cited problem with OPS is that OPS gives equal weight to its two components, OBP and SLG. In fact, OBP contributes significantly more to scoring runs than SLG does. This paper provides some exploration into the correct weighting of OBP to SLG when adding the two together. By correlating different coefficients of OBP to runs scored per game, the weighted OPS that weighting OBP 56% in two place more than SLG produced the highest correlation. We found that the weight of OBP increases as RPG increases. Also we suggest the linear regression equation of the best OBP coefficient against RPG.

Keywords: Correlation, OBP, OBP coefficient, OPS, sabermetrician, SLG.

¹ Professor, Department of Applied Statistics, Dankook University, Yongin 448-701, Korea.
E-mail: jtlee@dankook.ac.kr