

# Fielding indices for explaining runs lost combining adjusted WHIP and the number of home runs allowed in Korean professional baseball

Hyuk Joo Kim<sup>a,1</sup>

<sup>a</sup>Division of Mathematics & Informational Statistics, Wonkwang University

(Received August 16, 2016; Revised October 11, 2016; Accepted October 20, 2016)

---

## Abstract

We studied fielding indices to explain runs lost for Korean professional baseball teams, successively motivated by OPS and weighted OPS obtained by combining on-base percentage and slugging average that can adequately explain the run productivity of teams. We considered several combined indices made by combining fielding indices highly correlated with the runs lost of teams. Data analysis from all games in the regular seasons of 1982–2015 shows that weighted adjusted WPH 2 (defined as weighted average of adjusted WHIP and number of home runs allowed per inning) best explains runs lost. Weighted adjusted WPH 2 consisting of adjusted WHIP (with weight 34%) and number of home runs allowed per inning (with weight 66%) was found to be optimal weighted adjusted WPH 2 having correlation coefficient 0.95362 with average runs lost per game. This result is an improvement of the result of the index obtained in Kim and Kim (2015a). Analysis by chronological periods provides results that are not much different. Also we made a list of top 10 pitchers for each of the recent three years, based on the obtained index.

Keywords: adjusted WHIP, number of home runs allowed per game, number of home runs allowed per inning, weighted adjusted WPH, runs lost

---

## 1. 서론

많은 스포츠 종목은 승부를 가리는 것을 본질로 한다. 그중에서도 거의 모든 구기 종목 경기는 두 팀이 맞붙어 득점과 실점에 의해 승부를 결정한다(배구나 탁구처럼 세트제로 하는 종목도 있지만, 각 세트도 득점이 많아야 따고, 세트를 많이 따야 경기에서 이기는 것이다). 즉 경기에서 팀이 승리하려면 그 팀의 득점이 실점(상대팀의 득점)보다 많아야 한다. 야구도 마찬가지여서, 경기에서 안타나 홈런을 아무리 많이 쳐도 상대팀에 내준 점수가 한 점이라도 더 많으면 그 팀은 진다. 따라서 모든 팀은 득점을 늘리고 실점을 줄이기 위해 최선을 다하는 것이다.

한국 프로야구팀들의 득점과 공격 능력 및 실점에 관한 국내 학자들의 연구를 보면 다음과 같다. Cho와 Cho (2005a)는 on-base percentage plus slugging average(OPS)가 득점에 관한 매우 좋은 추정 도구가

---

This paper was supported by Wonkwang University in 2015.

<sup>1</sup>Division of Mathematics & Informational Statistics and Institute of Basic Natural Sciences, Wonkwang University, 460, Iksan-daero, Iksan, Jeonbuk 54538, Korea. E-mail: [hjkim@wonkwang.ac.kr](mailto:hjkim@wonkwang.ac.kr)

된다는 점을 밝혔으며, Cho와 Cho (2005b)는 득점과 실점을 이용하여 승률을 추정하는 문제를 연구하였다. Lee와 Kim (2005)은 타자들의 공격 능력의 측정 도구로 제안되는 여러 가지의 득점 측도를 비교하였고, Hong과 Park (2008)은 한국 프로야구에서 팀들이 기록한 안타의 방향을 분석하고 이를 시각적으로 표현하는 방법을 개발하였다. Hong과 Choi (2008)는 2007년 한국 프로야구에서의 도루 성공 모형을 연구하였으며, Chea 등 (2010)은 정규리그 자료를 이용하여 포스트시즌 진출 여부를 예측할 수 있는 여러 가지 통계모형의 분석 결과를 비교하였다. Lee (2014a)는 타자의 능력을 측정하는 지표를 제안하였으며, Lee (2014b)는 출루율 계수를 추정하는 문제를 연구하였다. 한편 Kim (2012)은 가중OPS를 제안하여 팀의 득점력을 설명하였으며, Kim과 Kim (2014)은 가중OPS를 수정한 가중수정OPS를 이용하면 팀의 득점력을 좀 더 잘 설명할 수 있음을 보였다.

실점과 관련이 많은 투수의 자책점에 관련된 연구로서 Cho와 Cho (2004)는 2003년 시즌의 한국 프로야구 자료를 바탕으로 walks plus hits divided by innings pitched(WHIP)가 평균자책점(earned run average; 방어를이라고도 함)에 미치는 영향을 연구하였다. WHIP는 최근 몇 년 전부터 투수의 능력을 평가하는 중요한 지표로 사용되고 있는 것으로서, 다음 식으로 정의된다.

$$\text{WHIP} = \frac{\text{볼넷} + \text{피안타}}{\text{투구이닝}}$$

이로부터 WHIP가 팀의 실점에도 영향을 미칠 것으로 추측할 수 있다. 하지만 출루율과 장타율이 결합된 OPS를 사용하면 출루율과 장타율 중 하나만을 사용할 때보다 팀 득점을 훨씬 더 잘 설명할 수 있듯이, 실점의 경우에도 두 개의 지표를 결합한 적절한 수비지표를 사용하면 하나의 지표만을 사용하는 경우보다 설명력이 높아질 것으로 기대할 수 있다. 이런 의미에서 Kim과 Kim (2015a)은 팀들의 실점을 설명하기 위한 수비지표를 연구하였다. 여러 가지의 수비지표 중 팀의 실점과 관련이 큰 것들을 결합하여 만든 몇 가지의 결합 지표들을 고려하였다. 프로야구 원년인 1982년부터 2014년까지의 정규리그 전 경기 자료를 분석한 결과 WHIP와 경기당피홈런의 가중평균으로 정의되는 가중WPH가 실점을 가장 잘 설명해줬다. 구체적으로 WHIP에 81%, 경기당피홈런에 19%의 가중치를 주는 가중WPH가 팀의 평균실점과 0.95033의 상관계수를 갖는 최적의 가중WPH인 것으로 나타났다. 한편 Kim과 Kim (2015b)은 변수선택 기법을 사용하여 팀의 득점과 실점을 회귀모형으로 설명하는 문제를 연구하였다.

그런데 WHIP를 정의하는 식을 보면 피안타와 볼넷만 포함되고 몸에 맞는 볼은 제외되어 있다. 주자를 내보내서 실점의 위기를 맞는다는 점에서는 볼넷과 몸에 맞는 볼이 차이가 없으므로 이를 반영하여 팀의 실점을 설명하는 데 사용하면 설명력이 좀 더 높아지지 않을까 기대된다. 따라서 본 논문에서는 수정된 WHIP와 기타의 수비지표들을 사용하여 팀들의 실점을 설명하기 위한 수비지표를 연구하고자 한다.

## 2. 몇 가지의 단일 수비지표와 실점 사이의 관계

본 논문에서 사용된 자료는 한국 프로야구 원년인 1982년부터 2015년까지 34년에 걸친 정규리그 전 경기(16,380경기)의 수비자료이며, 연 263개 팀의 1년 단위 기록들을 근거로 하였다. 이 자료는 Korea Baseball Organization (2009)의 한국프로야구 기록대백과와 Korea Baseball Organization (2010-2016)의 한국프로야구 연감에서 구한 것이다. 본 논문의 통계분석과 그래프 작성은 모두 미니탭 14.12.1을 이용하여 수행하였다.

여러 가지의 단일 수비지표와 실점 사이의 관계를 알아보았다. 팀들의 피안타율과 피출루율, 경기당 허용타석 등 여러 지표와 경기당 평균실점 간의 Pearson 상관계수(이하 ‘상관계수’로 표기)를 구하여 Table 2.1에 나타냈다. 또한 각각의 지표를 기준으로 하여 팀들의 순위를 구한 뒤 경기당 평균실점을 기준으로 한 순위와의 Spearman 순위상관계수(이하 ‘순위상관계수’로 표기)를 구하여 표에 함께 나타냈

**Table 2.1.** Correlation coefficient and rank correlation coefficient between each fielding index and average runs lost per game

	PAG	ABG	AVG	BHG	OBP	WHIP	AWHIP
Corr. coeff.	0.87152	0.73623	0.87619	0.60094	0.88520	0.91016	<b>0.92027</b>
Rank corr. coeff.	0.85997	0.69534	0.86069	0.60948	0.87963	0.90264	<b>0.91285</b>
	HR	HRG	HRI	SAG	KG	WPG	ERG
Corr. coeff.	0.73014	0.75963	0.76632	0.08901	0.20771	0.55305	-0.03647
Rank corr. coeff.	0.72485	0.73927	0.74554	0.08239	0.24393	0.53861	-0.11572

다. 상관계수는 실점에 대한 설명력이 높은 단일 수비지표를 찾기 위해 구한 것이며, 순위상관계수는 지푃값의 순위 사이의 상관계수가 지푃값 자체로 구한 상관계수와 얼마나 유사한지를 보기 위해 참고로 구한 것이다.

위의 표에서 PAG는 경기당허용타석, ABG는 경기당허용타수, AVG는 피안타율, BHG는 경기당4사구(여기서 4사구는 볼넷과 몸에 맞는 볼을 통칭한 것이다), OBP는 피출루율, HR은 피홈런, HRG는 경기당피홈런, HRI는 이닝당피홈런, SAG는 경기당희생타, KG는 경기당탈삼진, WPG는 경기당폭투, ERG는 경기당실책을 각각 나타낸다. 또한 AWHIP는 다음 식으로 정의되는 수정WHIP(adjusted WHIP)를 나타내는 것이다.

$$\text{수정WHIP} = \frac{4\text{사구} + \text{피안타}}{\text{투구이닝}}$$

위의 표에서 볼 수 있듯이 상관계수와 순위상관계수는 거의 모든 지표에서 매우 유사한 값을 갖는다. 여러 지표들 중 평균실점과 높은 상관관계를 갖는 것을 순서대로 써보면 수정WHIP(상관계수 0.92027), WHIP(상관계수 0.91016), 피출루율(상관계수 0.88520), 피안타율(상관계수 0.87619), 경기당허용타석(상관계수 0.87152), 이닝당피홈런(상관계수 0.76632), 경기당피홈런(상관계수 0.75963)이다. 그리고 피홈런수 자체보다는 피홈런수를 이닝수로 나눈 이닝당피홈런과 경기수로 나눈 경기당피홈런이 평균실점과 좀 더 높은 상관관계를 가지며, 탈삼진과 실책은 평균실점과의 상관관계가 약한 것으로 나타났다. 투구이닝과 보크, 야수의 수비율 등에 관해서도 조사했는데, 평균실점과의 상관관계가 거의 없는 것으로 나왔기 때문에 표에 나타내지 않았다.

### 3. 실점을 설명하기 위한 결합 수비지표

출루율과 장타율의 단순합계인 OPS는 현재 중요한 공격지표로 사용되고 있다. 출루율과 장타율 중 어느 한 쪽에 더 큰 비중을 둔 가중평균을 사용한다면 팀의 득점력을 좀 더 잘 설명할 수 있을 것인데, 두 가지를 똑같은 비중으로 결합한 OPS가 널리 사용되고 있는 것은 단지 계산이 간편하기 때문이다. Kim (2012)은

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

로 정의되는 가중OPS를 제안하여 득점력을 좀 더 잘 설명하고자 하였다. 여기서  $a$ 는 0과 1 사이의 값이다. 2007년부터 2011년까지의 한국 프로야구 정규리그 전 경기 자료를 분석한 결과 출루율에 57%의 비중을 두고 장타율에 43%의 비중을 두는 가중OPS가 팀의 득점력을 가장 잘 설명하는 것으로 밝혀졌다. 한편 Kim과 Kim (2014)은 출루율을 약간 수정하여 정의되는 수정출루율과 장타율의 가중평균인 가중수정OPS를 이용하면 팀의 득점력을 좀 더 잘 설명할 수 있음을 보였다.

본 논문에서는 중요한 공격지표 두 개(출루율과 장타율)의 가중평균으로 정의되어 팀의 득점력을 잘 설명해주는 가중OPS처럼, 중요한 수비지표 두 개의 가중평균으로 정의되어 팀의 실점을 잘 설명해주는 수비지표를 제시하고자 한다. 2절에서 평균실점과 높은 상관관계를 갖는 지표들이 수정WHIP, WHIP, 피출루율, 피안타율, 경기당허용타석, 이닝당피홈런, 경기당피홈런으로 밝혀졌으므로, 이것들을 두 개씩 짝지어 결합하여 다음과 같이 19가지의 결합 수비지표를 고려한다. WHIP와 수정WHIP의 결합, 그리고 경기당피홈런과 이닝당피홈런의 결합은 의미가 없으므로 제외하였다. 지표들의 이름은 기존의 용어인 ‘OPS’가 만들어진 방식대로 편의상 영문약자를 만들어 붙여본 것이다. 예를 들어 ‘PPA’는 ‘타석(plate appearance)’의 ‘P’와 ‘Plus’의 ‘P’, 그리고 ‘타율(average)’의 ‘A’를 따서 만든 것이다. 아래의 19가지 중 WHIP와 경기당피홈런을 결합한 가중WPH1은 Kim과 Kim (2015a)에서 제안된 것이다.

$$\begin{aligned} \text{가중PPA} &= a \times \text{경기당허용타석} + (1 - a) \times \text{피안타율} \\ \text{가중PPO} &= a \times \text{경기당허용타석} + (1 - a) \times \text{피출루율} \\ \text{가중PPW} &= a \times \text{경기당허용타석} + (1 - a) \times \text{WHIP} \\ \text{가중수정PPW} &= a \times \text{경기당허용타석} + (1 - a) \times \text{수정WHIP} \\ \text{가중PPH1} &= a \times \text{경기당허용타석} + (1 - a) \times \text{경기당피홈런} \\ \text{가중PPH2} &= a \times \text{경기당허용타석} + (1 - a) \times \text{이닝당피홈런} \\ \text{가중APO} &= a \times \text{피안타율} + (1 - a) \times \text{피출루율} \\ \text{가중APW} &= a \times \text{피안타율} + (1 - a) \times \text{WHIP} \\ \text{가중수정APW} &= a \times \text{피안타율} + (1 - a) \times \text{수정WHIP} \\ \text{가중APH1} &= a \times \text{피안타율} + (1 - a) \times \text{경기당피홈런} \\ \text{가중APH2} &= a \times \text{피안타율} + (1 - a) \times \text{이닝당피홈런} \\ \text{가중OPW} &= a \times \text{피출루율} + (1 - a) \times \text{WHIP} \\ \text{가중수정OPW} &= a \times \text{피출루율} + (1 - a) \times \text{수정WHIP} \\ \text{가중OPH1} &= a \times \text{피출루율} + (1 - a) \times \text{경기당피홈런} \\ \text{가중OPH2} &= a \times \text{피출루율} + (1 - a) \times \text{이닝당피홈런} \\ \text{가중WPH1} &= a \times \text{WHIP} + (1 - a) \times \text{경기당피홈런} \\ \text{가중WPH2} &= a \times \text{WHIP} + (1 - a) \times \text{이닝당피홈런} \\ \text{가중수정WPH1} &= a \times \text{수정WHIP} + (1 - a) \times \text{경기당피홈런} \\ \text{가중수정WPH2} &= a \times \text{수정WHIP} + (1 - a) \times \text{이닝당피홈런} \end{aligned}$$

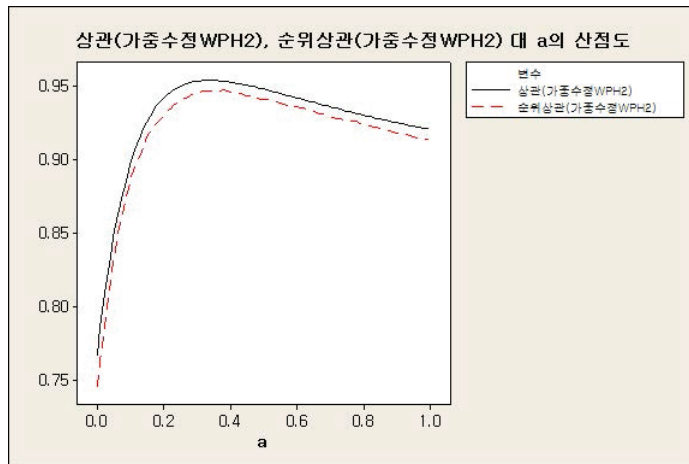
위의 식에서  $a$ 는 0과 1 사이(0과 1 포함)의 값이다.  $a$ 의 값을 0부터 1까지 0.01 간격으로 변화시켜 가며 팀들의 수비지표들을 계산하고, 이 수비지표와 팀의 경기당 평균실점 간의 상관계수를 구하여, 각 수비지표에 대해 상관계수가 최대가 되는  $a$ 값과 그때의 상관계수를 구한 결과가 Table 3.1에 나와 있다. 이 표에서 각 지표의 이름 중 맨 앞의 W는 ‘가중’을 의미한다.

위 표에 있는 값들 중 가장 큰 값은  $a = 0.34$ 일 때의 가중수정WPH2와 평균실점 간의 상관계수인 0.95362이다. 이것은 가중평균을 이용한 지표 중에서 다음 식과 같이 수정WHIP에 34%, 이닝당피홈런에 66%의 비중을 두는 가중수정WPH2를 사용할 때 팀들의 실점이 가장 잘 설명된다는 것을 의미한다.

$$\text{가중수정WPH2}(0.34) = 0.34 \times \text{수정WHIP} + 0.66 \times \text{이닝당피홈런}.$$

**Table 3.1.** Maximum correlation coefficient (with  $a$  value) between each combined fielding index and average runs lost per game

Fielding index	WPPA	WPPO	WPPW	WAPPW	WPPH1
$a$	0.02	0.02	0.04	0.02	0.31
Max. corr. coeff.	0.91768	0.90211	0.91533	0.92183	0.91965
Fielding index	WPPH2	WAPO	WAPW	WAAPW	WAPH1
$a$	0.05	0.49	0.76	0.74	0.96
Max. corr. coeff.	0.92257	0.90490	0.91857	0.92727	0.93167
Fielding index	WAPH2	WOPW	WAOPW	WOPH1	WOPH2
$a$	0.74	0.00	0.00	0.96	0.71
Max. corr. coeff.	0.93227	0.91016	0.92027	0.94219	0.94285
Fielding index	WWPH1	WWPH2	WAWPH1	WAWPH2	
$a$	0.81	0.32	<b>0.82</b>	<b>0.34</b>	
Max. corr. coeff.	0.95261	0.95290	<b>0.95340</b>	<b>0.95362</b>	



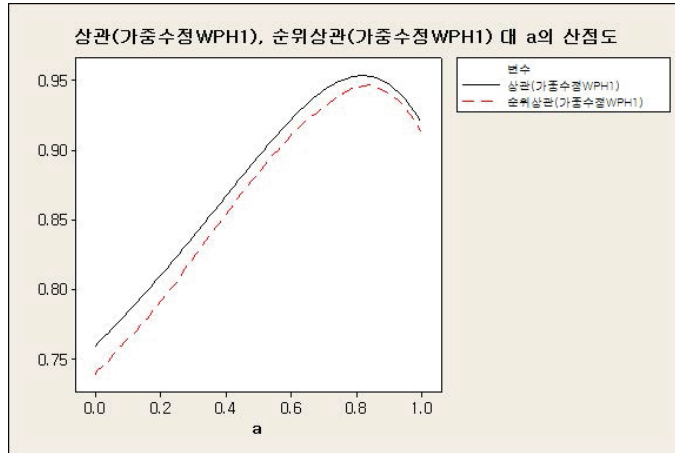
**Figure 3.1.** Plot of correlation coefficient and rank correlation coefficient between weighted adjusted WPH2 and average runs lost per game, against  $a$ .

또한 두 번째로 큰 값은  $a = 0.82$ 일 때 가중수정WPH1과 평균실점 간의 상관계수인 0.95340이다. 이것은 다음과 같이 수정WHIP에 82%, 경기당피홈런에 18%의 비중을 두는 가중수정WPH1을 의미한다.

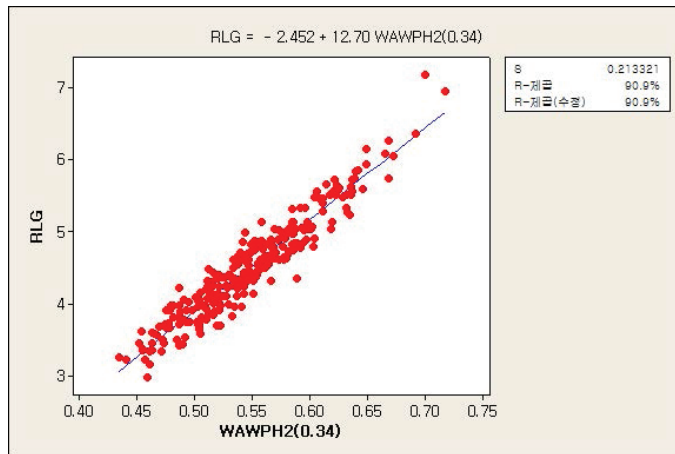
$$\text{가중수정WPH1}(0.82) = 0.82 \times \text{수정WHIP} + 0.18 \times \text{경기당피홈런},$$

여기서 알 수 있듯이 가중수정WPH1과 가중수정WPH2의 최적의  $a$ 값 간에 큰 차이가 있다. 수정WHIP를 경기당피홈런과 결합할 때는 수정WHIP의 비중이 82%일 때가 최적이지만, 이닝당피홈런과 결합할 때는 수정WHIP의 비중이 34%일 때가 최적인 것으로 나타났다.

가중수정WPH2와 평균실점 간의 상관계수와 순위상관계수가  $a$ 값에 따라 변화하는 내용을 그래프로 작성한 것이 Figure 3.1이다. 이 그림에서 실선은 상관계수를 나타내고 점선은 순위상관계수를 나타낸다. 상관계수는  $a$ 값이 0에서 0.34가 될 때까지 급격히 증가하다가 0.34를 지나면서 1이 될 때까지는 서서히 감소하는 모양을 보인다. 순위상관계수는  $a = 0.38$ 일 때 최댓값 0.94704를 갖는 것으로 밝혀졌고, 변화



**Figure 3.2.** Plot of correlation coefficient and rank correlation coefficient between weighted adjusted WPH1 and average runs lost per game, against  $a$ .



**Figure 3.3.** Plot of average runs lost per game against WAWPH2(0.34)(1982–2015).

하는 패턴은 상관계수의 경우와 유사하다. Figure 3.2는 가중수정WPH1과 평균실점 간의 상관계수와 순위상관계수가  $a$ 값에 따라 변화하는 내용을 그래프로 작성한 것이다. Figure 3.1과 Figure 3.2에서 보 다시피 수정WHIP와 결합하는 수비지표로 이닝당피홈런을 사용할 때와 경기당피홈런을 사용할 때가 상당히 다른 양상을 보인다.

가중수정WPH2(0.34)가 팀의 경기당 평균실점에 관한 좋은 지표가 되려면 가중수정WPH2(0.34)와 경기당 평균실점 사이에 직선에 가까운 유의한 관계가 존재해야 할 것이다. 이를 확인하기 위하여 이 두 변수 간의 산점도를 최소제곱법에 의한 회귀직선과 함께 나타낸 것이 Figure 3.3이다. 이 그림에서 RLG는 경기당 평균실점을 의미한다. 회귀분석의 결과를 Table 3.2와 Table 3.3에 나타냈다. 가중수정WPH2(0.34)를  $x$ , 경기당 평균실점을  $y$ 라 할 때, 최소제곱법에 의한 회귀직선의 방정식은

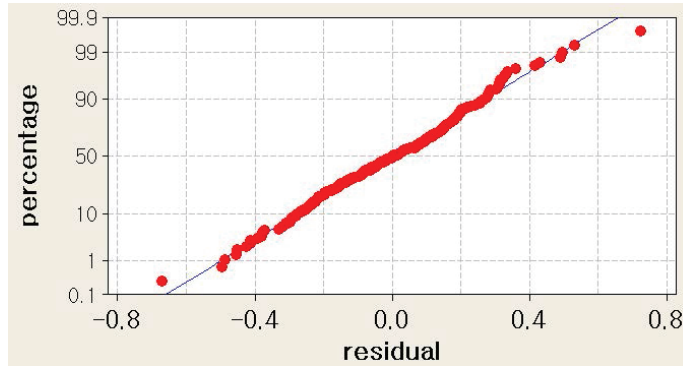
$$\hat{y} = -2.4525 + 12.7017x$$

**Table 3.2.** Result of regression analysis

Variable	Estimate	Standard error	<i>t</i> -value	<i>p</i> -value	<i>r</i> <sup>2</sup>
Intercept	-2.4525	0.1367	-17.94	< 0.001	
WWPH2(0.34)	12.7017	0.2482	51.18	< 0.001	0.9094

**Table 3.3.** Analysis of variance for the regression

Source	Sum of squares	Degrees of freedom	Mean square	<i>F</i> -value	<i>p</i> -value
Regression	119.2131	1	119.2131	2619.74	< 0.001
Residual	11.8770	261	0.0455		
Total	131.0901	262			



**Figure 3.4.** Normal probability plot of residuals for the fitted regression model.

**Table 3.4.** Unusual observations based on the regression equation

Year	Team	WAWPH2 (0.34)	RLG	$\widehat{RLG}$	$\widehat{SE}(\widehat{RLG})$	Residual	Standardized residual
1982	Sammi	0.701	7.1750	6.4501	0.0401	0.7249	3.46
1988	Pacific	0.543	4.9815	4.4491	0.0132	0.5324	2.50
1998	Lotte	0.558	5.1270	4.6311	0.0134	0.4959	2.33
1996	Hanwha	0.487	4.2222	3.7327	0.0201	0.4895	2.30
1982	MBC	0.512	4.4875	4.0555	0.0159	0.4320	2.03
2003	LG	0.589	4.3534	5.0239	0.0165	-0.6705	-3.15
2005	Doosan	0.523	3.6905	4.1874	0.0146	-0.4970	-2.34
2007	Hanwha	0.532	3.8175	4.3054	0.0138	-0.4880	-2.29
2007	SK	0.519	3.6905	4.1447	0.0150	-0.4542	-2.13
2003	KIA	0.541	3.9624	4.4138	0.0133	-0.4514	-2.12
2004	LG	0.604	4.7895	5.2148	0.0190	-0.4253	-2.00

었다. Figure 3.4는 위의 적합된 모형에 따른 잔차의 정규확률그림(normal probability plot)을 그린 것이다. 오차항의 정규성 가정에 별다른 문제가 없음을 알 수 있다. 모형의 결정계수는 90.94%(이 값은 물론 앞에서 얻은 상관계수 0.95362의 제곱과 같다)로 매우 높았으며, 회귀분석의 *F*-통계량의 값은 2,619.74였고 *p*-값은 0.001보다 작았다. Table 3.4는 위의 회귀방정식에 근거를 두고 표준화잔차의 절댓값이 큰 비정상적인 관측값들을 찾아낸 것이다. 예컨대 1982년에 삼미의 실제 평균실점 7.1750점으로 가중수정WPH2(0.34)에 의한 적합치 6.4501점보다 비정상적으로 많이 나왔다는 것을 의미하며,

**Table 4.1.** Correlation coefficients between each single fielding index and average runs lost per game (by chronological periods)

	WHIP	AWHIP	HRG	HRI
1982–1989	0.89842	0.91375	0.75701	0.76723
1990–1999	0.90235	0.91456	0.76170	0.76556
2000–2009	0.87590	0.89154	0.72935	0.73941
2010–2015	0.93350	0.93429	0.75423	0.79454

**Table 4.2.** Four combined fielding indices highly correlated with average runs lost per game (by chronological periods)

	First	Second	Third	Fourth
1982–1989	WAWPH2	WAWPH1	WWPH2	WWPH1
	0.95150 ( $a = 0.27$ )	0.95076 ( $a = 0.77$ )	0.94864 ( $a = 0.25$ )	0.94771 ( $a = 0.75$ )
1990–1999	WAWPH1	WAWPH2	WWPH2	WWPH1
	0.95712 ( $a = 0.81$ )	0.95704 ( $a = 0.32$ )	0.95348 ( $a = 0.31$ )	0.95346 ( $a = 0.80$ )
2000–2009	WAWPH1	WAWPH2	WWPH1	WWPH2
	0.94762 ( $a = 0.82$ )	0.94720 ( $a = 0.33$ )	0.94617 ( $a = 0.80$ )	0.94586 ( $a = 0.31$ )
2010–2015	WAWPH2	WAWPH1	WWPH2	WWPH1
	0.96783 ( $a = 0.34$ )	0.96758 ( $a = 0.82$ )	0.96654 ( $a = 0.35$ )	0.96630 ( $a = 0.83$ )

1988의 태평양, 1998년의 롯데, 1996년의 한화, 1982년의 MBC도 유사한 경우이다. 반면 2003년의 LG, 2005년의 두산, 2007년의 한화와 SK, 2003년의 KIA, 2004년의 LG는 실제 평균실점이 가중수정WPH2(0.34)에 의한 적합치보다 비정상적으로 적게 나온 경우이다.

#### 4. 시대별 분석

이 절에서는 한국 프로야구를 시대별(1980년대, 1990년대, 2000년대, 2010년대)로 구분하여 실점과 수비지표들의 관계를 살펴보고자 한다. Table 4.1에는 4개의 단일 수비지표 WHIP, 수정WHIP(AWHIP), 경기당피홈런(HRG), 이닝당피홈런(HRI)과 경기당 평균실점 사이의 상관계수를 시대별로 구한 것이 나와 있다. 모든 시대에서 공통적으로 수정WHIP가 WHIP보다, 그리고 이닝당피홈런이 경기당피홈런보다 조금씩 더 높은 상관관계를 보이고 있다.

한편 3절에서 고려한 19개의 결합 수비지표들에 대해 시대별로 경기당 평균실점과의 상관계수가 최대가 되는  $a$ 값과 그때의 상관계수를 구한 뒤 이 최대상관계수의 값이 큰 순으로 4개를 나타낸 것이 Table 4.2에 나와 있다. 표에서 보는 바와 같이 모든 시대에 걸쳐 공통적으로 가중수정WPH(1과 2)가 1–2위를 차지하고 있으며, 가중WPH(1과 2)가 3–4위이다. 이것은 각 시대별로도 기존의 WHIP가 반영된 결합 수비지표보다 수정WHIP가 반영된 결합 수비지표가 경기당 평균실점을 더욱 잘 설명한다는 것을 의미한다. 그리고 1980년대와 2010년대에는 이닝당피홈런을 사용한 가중수정WPH2가 경기당피홈런을 사용한 가중수정WPH1보다 좀 더 설명력이 높게 나왔고, 1990년대와 2000년대에는 그 반대였다. 종합적으로는 3절에서 본 것과 같이 가중수정WPH2의 설명력이 좀 더 높았다. 상관계수를 최대로 하는  $a$ 값을 시대별로 보면, 가중수정WPH2의 경우 0.27, 0.32, 0.33, 0.34로 전 기간에 걸친 최적값 0.34와 큰 차이가 없었고 1980년대에 비해 1990년대 이후에 수정WHIP의 비중이 다소 커졌다. 가중수정WPH1의 경우도 시대별 최적  $a$ 값이 0.77, 0.81, 0.82, 0.82로 전 기간에 걸친 최적값 0.82와 큰 차이가 없었고 1980년대에 비해 1990년대 이후에 WHIP의 비중이 다소 커졌다. 그리고 최대상관계수의 값은 가중수정WPH2의 경우 2000년대에 약간 작아졌다가 2010년대에 가장 커졌으며, 가중수정WPH1의



**Table 5.1.** Top 10 pitchers based on WAWPH2(0.34) for each year (2013–2015)

Year	Rank	Pitcher	Team	AWHIP	HRI	WAWPH2(0.34)	ERA
2013	1	Charlie	NC	1.3175	0.0317	0.4689	2.48 (1)
	2	Liz	LG	1.2878	0.0543	0.4737	3.06 (4)
	3	YoonSH	Samsung	1.2188	0.0938	0.4763	3.27 (5)
	4	LeeJH	NC	1.2564	0.0769	0.4779	2.88 (2)
	5	Hacker	NC	1.2897	0.0673	0.4829	3.63 (9)
	6	Oxspring	Lotte	1.3309	0.0545	0.4885	3.29 (6)
	7	VandenHurk	Samsung	1.2947	0.0766	0.4907	3.95 (15)
	8	WooKM	LG	1.3914	0.0339	0.4955	3.91 (14)
	9	YooHK	Doosan	1.3761	0.0482	0.4997	3.53 (7)
	10	Seddon	SK	1.3505	0.0747	0.5085	2.98 (3)
2014	1	VandenHurk	Samsung	1.1463	0.0983	0.4546	3.18 (1)
	2	Van Hekken	Nexen	1.3316	0.0481	0.4845	3.51 (3)
	3	Riordan	LG	1.3333	0.0774	0.5044	3.96 (6)
	4	Nippert	Doosan	1.3383	0.0948	0.5176	3.81 (5)
	5	WooKM	LG	1.4056	0.0716	0.5252	4.04 (8)
	6	YangHJ	KIA	1.4183	0.0700	0.5284	4.25 (12)
	7	Hacker	NC	1.4189	0.0811	0.5359	4.01 (7)
	8	Oxspring	Lotte	1.3671	0.1085	0.5364	4.20 (10)
	9	Martin	Samsung	1.4531	0.0781	0.5456	4.78 (16)
	10	YoonSH	Samsung	1.4149	0.1057	0.5508	4.39 (13)
2015	1	Hacker	NC	1.1520	0.0686	0.4370	3.13 (2)
	2	Sosa	LG	1.2350	0.0823	0.4742	4.03 (11)
	3	WooKM	LG	1.2576	0.0852	0.4838	3.42 (4)
	4	Van Hekken	Nexen	1.3119	0.0712	0.4930	3.62 (6)
	5	YangHJ	KIA	1.2749	0.0976	0.4979	2.44 (1)
	6	Figaro	Samsung	1.2848	0.0970	0.5008	3.38 (3)
	7	Lindblom	Lotte	1.2381	0.1333	0.5090	3.56 (5)
	8	YoonSH	Samsung	1.2320	0.1392	0.5107	3.76 (8)
	9	YooHK	Doosan	1.2707	0.1213	0.5121	3.94 (10)
	10	Kelly	SK	1.3646	0.0884	0.5223	4.13 (13)

경우도 유사한 패턴을 보였다.

## 5. 투수 평가에의 적용

WHIP와 투수들의 평가에 사용되고 있는 대표적인 지표가 평균자책점이다. 이것은 투수가 허용한 자책점을 투구이닝 수로 나눈 뒤 9를 곱한 것으로, 한 경기(9회)당 평균적으로 허용한 자책점 수를 나타낸다.

평균자책점은 투수의 능력을 나타내는 객관적인 지표로 쓰이고 있으나 생각해보면 여기에도 약간의 불합리한 경우가 나올 수 있다. 예를 들어 투수 A가 안타와 볼넷 등으로 3명의 주자를 내보내 만루를 만들어 놓고 투수 B로 교체되었다고 하자. 이어 던진 투수 B가 잘 막아서 점수를 안 주면 투수 A의 자책점은 0점이 되지만, 만약 투수 B가 주자들의 득점을 허용하면 투수 A의 자책점은 3점이 된다. 즉 평균자책점은 자신이 아닌 타인의 투구 결과에 따라 큰 영향을 받는 경우가 흔히 있는 것이다.

위와 같은 점을 고려하면 본 논문에서 제시된 가중수정WPH2(0.34)를 사용하여 투수들을 평가하는 것도 의미가 있을 것으로 본다. Table 5.1은 2013년부터 2015년까지 연도별로 가중수정WPH2(0.34)을

지표로 하여 규정이닝 이상을 던진 투수들의 순위를 1위부터 10위까지 나타낸 것이다. 참고로 한국 프로야구 1부 리그에서 투수의 규정이닝 수는 소속팀의 경기 수와 같다. 지표의 값을 보면 2013년에는 NC의 찰리가 0.4689, 2014년에는 삼성의 벤덴헨크가 0.4546, 2015년에는 NC의 해커가 0.4370으로 각각 1위인 것으로 나타났다. 표의 맨 오른쪽에 ERA라고 표시된 것은 평균자책점이며, 괄호 안의 숫자는 평균자책점 순위이다. 가중수정WPH(0.34) 순위와 평균자책점 순위 사이에 다소 차이가 있는 것을 볼 수 있다.

## 6. 결론

최근 들어 투수의 능력을 나타내는 유용한 지표로 사용되고 있는 것이 WHIP이다. 이것은 투수의 피안타 수와 볼넷의 수를 합쳐서 투구이닝 수로 나눈 것이다. 본 논문에서는 WHIP에 몸에 맞는 볼의 수가 반영되지 않는다는 점에 주목하여 몸에 맞는 볼의 수가 반영된 지표를 고려하고 이를 수정WHIP라 칭하였다. 이 지표는 기존의 WHIP에 문제가 있으니 이걸로 교체해서 사용하자는 뜻으로 제시한 것은 아니고, 다만 팀의 실점과의 관련성이 조금 더 높으므로 실점을 설명하기 위한 결합 수비지표(가중수정WPH1과 가중수정WPH2)를 만들기 위한 용도로 제시한 것이다. 이 수정WHIP는 역대 한국 프로야구의 34개 시즌 중 21개의 시즌에서 기존의 WHIP보다 팀들의 실점을 더 잘 설명하였다. 또한 팀의 경기당 평균실점과 높은 상관관계를 갖는 7개의 단일 수비지표인 수정WHIP, WHIP, 피출루율, 피안타율, 경기당허용타석, 이닝당피홈런, 경기당피홈런을 두 개씩 짝지어 생각할 수 있는 19가지의 결합 수비지표(결합이 무의미한 두 가지는 제외) 중 평균실점을 가장 잘 설명해주는 수비지표인 가중수정WPH1과 가중수정WPH2를 얻었다. 전자는 수정WHIP와 경기당피홈런의 가중평균으로 정의된 것이며, 후자는 수정WHIP와 이닝당피홈런의 가중평균으로 정의된 것이다. 그중에서도 가중수정WPH2가 가중수정WPH1보다 좀 더 높은 설명력을 보였다. 한국 프로야구 원년인 1982년부터 2015년까지의 정규리그 전 경기 자료를 분석한 결과 수정WHIP에 34%, 이닝당피홈런에 66%의 비중을 두는 가중수정WPH2가 팀의 실점을 가장 잘 설명하는 것으로 나타났으며, 경기당 평균실점과의 상관계수는 0.95362였다.

본 논문에서 얻어진 가중수정WPH2는 Kim과 Kim (2015a)에서 제안된 가중WPH(3절에서는 가중WPH1으로 표시되었음)와 비교할 때 다음과 같은 장점을 갖는다. 우선 팀의 실점에 대한 설명력이 가중WPH보다 좀 더 높게 나타났다. 다음으로, 가중WPH는 경기당피홈런이 결합된 지표이기 때문에 개인기록으로는 사용될 수 없다는 한계를 갖는 데 반해, 가중수정WPH2는 개인의 평가에도 사용될 수 있다. 이 지표의 효용성은 추후에 얻어질 데이터를 통해서도 계속 검토할 수 있을 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. (2004). Study about the influence that WHIP has on ERA in 2003 season Korean professional baseball, *Journal of the Korean Data Analysis Society*, **6**, 1415–1424.
- 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 Choi, J. M. (2008). Steal success model for 2007 Korean professional baseball games, *The Korean Journal of Applied Statistics*, **21**, 455–468.

- Hong, C. S. and Park, H. S. (2008). Visual representation and applications of hitting direction in Korean baseball records, *Journal of the Korean Data and 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 and Information Science Society*, **23**, 1065–1074.
- Kim, H. J. and Kim, Y. H. (2014). Explanation of run productivity using weighted adjusted OPS in Korean professional baseball, *The Korean Journal of Applied Statistics*, **27**, 731–741.
- Kim, H. J. and Kim, Y. H. (2015a). Explanation of runs lost using combined fielding indices in Korean professional baseball, *The Korean Journal of Applied Statistics*, **28**, 1003–1011.
- Kim, H. J. and Kim, Y. H. (2015b). Explanation of runs using variable selection technique in Korean professional baseball, *Journal of the Korean Official Statistics*, **20**, 71–92.
- Korea Baseball Organization (2009). *The Official Baseball Encyclopedia 2009*, Korea Baseball Organization, Seoul.
- Korea Baseball Organization (2010–2016). *2010–2016 Official Baseball Guide*, Korea Baseball Organization, Seoul.
- 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 and 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 수정된 WHIP와 피홈런 수를 결합한 실점 설명 수비지표들

김혁주<sup>a,1</sup>

<sup>a</sup>원광대학교 수학·정보통계학부

(2016년 8월 16일 접수, 2016년 10월 11일 수정, 2016년 10월 20일 채택)

---

## 요약

한국 프로야구팀들의 실점을 설명하기 위한 수비지표를 연구하였다. 출루율과 장타율을 결합한 OPS가 팀들의 득점력을 매우 잘 설명한다는 사실을 동기로 하여 본 논문에서는 팀들의 실점을 설명하기 위한 수비지표를 연구하였다. 여러 가지의 수비지표 중 팀의 실점과 상관관계가 높은 것들을 결합하여 만든 몇 가지의 결합 지표들을 고려했다. 프로야구 원년인 1982년부터 2015년까지의 정규리그 전 경기 자료를 분석한 결과 수정WHIP와 이닝당피홈런의 가중평균으로 정의되는 가중수정WPH2가 실점을 가장 잘 설명해줬다. 구체적으로 수정WHIP에 34%, 이닝당피홈런에 66%의 가중값을 두는 가중수정WPH2가 팀의 경기당 평균실점과 0.95362의 상관계수를 갖는 최적의 지표인 것으로 나타났다. 이 결과는 Kim과 Kim (2015a)에서 얻어진 지표에 의한 결과보다 좀 더 향상된 결과이다. 시대별 분석에서도 크게 다르지 않은 결과를 얻었다. 또한 얻어진 지표의 값이 좋은 투수 10명씩의 명단을 최근 3년간에 대해 연도별로 작성해보았다.

주요용어: 수정WHIP, 경기당피홈런, 이닝당피홈런, 가중수정WPH, 실점

---

이 논문은 2015학년도 원광대학교의 교비 지원에 의해서 수행됨.

<sup>1</sup>(54538) 전북 익산시 익산대로 460, 원광대학교 수학·정보통계학부 및 기초자연과학연구소.

E-mail: hjkim@wonkwang.ac.kr