

2003시즌 한국프로야구에서 WHIP가 방어율에 미치는 영향에 관한 연구

조영석¹ · 조용주²

요 약

투수의 능력을 파악하는데 가장 보편적으로 사용되는 것이 승률, 다승, 그리고 방어율을 사용합니다. 그리고 구원투수의 경우는 홀드, 세이브가 주로 사용되고 있습니다. 이는 KBO(한국야구위원회)지정 공식 타이틀 이기도 합니다. 그리고 투수의 능력을 측정하는데는 최근에 이닝당 삼진, 볼넷과 삼진의 비율, 피안타율을 많이 사용합니다. 방어율은 투수의 능력을 파악하는 가장 보편 타당한 수치이다. 왜냐하면 방어율은 투수가 타자와의 대결에서 만들어지는 수치이기 때문이다. 본 연구에서는 이러한 방어율을 추정하기 위하여 WHIP(walk+hit inning pitch)의 개념과 이용실태를 소개하고, 2003년 한국프로야구자료를 이닝 당 안타(inning hit : IH)와 이닝 당 볼넷(inning walk : IW)을 이용하여 군집분석을 실시하여 어떤 성격의 군집이 방어율이 가장 좋은지 알아보고 투수의 보직에 따라서 WHIP를 비교하였다. 그리고 IH와 IW를 이용하여 방어율에 대한 회귀식을 유도하고, WHIP를 이용한 회귀식과 IH와 IW를 이용한 회귀식을 비교하였다.

주요용어 : 방어율, WHIP, 군집분석, 회귀분석.

1. 서론

ESPN의 유명 칼럼니스트 Roy Never (2002)는 자신의 페이지를 통해 타자의 홈런과 볼넷, 투수의 피홈런과 피볼넷의 랭킹을 합산한 포인트 Beane Count(BC)라는 새로운 팀 랭킹을 제공하고 있다. 조영석과 조용주(2003)는 2002년 한국프로야구자료(사단법인 한국야구위원회, 2003)를 BC와 팀 승률간의 상관관계를 조사하고, 타자의 홈런과 볼넷, 투수의 피홈런과 피볼넷 등 4개의 변수를 군집분석 하여 어떠한 성격의 군집이 승률이 높은가를 판단하고, 회귀분석을 통하여 승률을 예측하는 모형을 제시하였다.

ESPN 홈페이지를 방문해보면 스포츠에 대한 기록을 정말 잘 정리해 놓은 것을 알 수 있다. 개인적으로 한국에는 이러한 사이트가 아직 없다는 것이 안타까운 생각이 든다. 특히 투수들의 기록을 보면 우리나라에서는 아직 잘 사용하고 있지 않는 WHIP(walk+hit inning pitch)라는 척도가 있다. WHIP는 한 이닝에 투수가 허용하는 안타와 볼넷의 개수의 합을 나타낸 것이다. 투수를 측정하는

¹(교신저자) 627-702 경상남도 밀양시 내이동 1025-1, 밀양대학교 자율전공학부 조교수

E-mail : choys@mmu.ac.kr

²712-749 경상북도 경산시 대동 214-1, 영남대학교 통계학과 석사. E-mail : nuriya@empal.com

가장 중요한 요소 가운데 하나가 바로 방어율이다. 승수나 승률은 타자의 도움으로도 기록할 수 있지만 방어율은 오직 투수 스스로의 힘으로 만들어 내는 것이기 때문이다. 방어율을 계산할 때 팀의 실점이 아닌 자기 자신의 자책점으로 나타내는 것을 보아도 알 수 있다. WHIP는 메이저리그에서도 공식적으로 산출하는 기록은 아니다. 그러나 투수의 방어율을 결정한다는 점에서 최근 매우 중요하게 여기는 항목이다. 스포츠 서울의 전문기자 칼럼에서 문상열 기자는 “미국의 야구 기록 분석가들에 따르면 WHIP와 방어율은 상호 밀접한 관련이 있으며 WHIP가 0.99가 되면 방어율은 2.06을 기록한다. 이는 3년간의 통계기록이다”라고 서술하고 있다(스포츠서울 전문기자 칼럼, 2003).

본 논문에서는 2003한국프로야구자료를 이용하여 WHIP와 방어율의 상관관계를 알아보고, 회귀 분석을 실시하고, IW와 IH를 군집분석을 실시하여 어떤 군집의 유형이 가장 방어율이 높은지 알아보고, 회귀모형을 구축하고자 한다. 그리고 방어율을 예측하기 위한 어떤 회귀모형이 설명력이 더 좋은지 알아보하고자 한다.

2. WHIP의 개념과 이용실태

투수의 능력을 측정하는 중요한 도구인 방어율이 결과적인 수치라면 WHIP는 원인을 제공하는 수치라고 할 수 있다. 즉 투수가 실점할 수 있는 가능성에 대한 수치라고 할 수 있다. 한 이닝에 투수의 능력으로 즉 볼넷이나 안타를 허용하는 비율이 낮으면 낮을수록 투수는 실점을 할 확률이 작아지는 것이다. 아래 [표 1]은 ESPN 홈페이지에서 2003시즌 뉴욕 메츠 투수들의 성적을 나타낸 것이다. 이닝 수가 많은 순서로 투수를 나열한 것인데 서재웅 선수는 팀내에서 2번째로 많은 이닝을 던졌고, WHIP는 1.27로 가장 낮게 나타났다.

[표 1] 2003 시즌 뉴욕 메츠 투수들의 성적

Pitching Statistics																
NAME	G	GS	W	L	Sv	Hld	IP	H	ER	HR	BB	SO	K/9	P/GS	WHIP	ERA
S. Trachsel	33	33	16	10	0	0	204.2	204	86	26	65	111	4.88	100.7	1.31	3.78
Jae Weong Seo	32	31	9	12	0	0	188.1	193	80	18	46	110	5.26	92.4	1.27	3.82
Tom Glavine	32	32	9	14	0	0	183.1	205	92	21	66	82	4.03	91.5	1.48	4.52
Al Leiter	30	30	15	9	0	0	180.2	176	80	15	94	139	6.92	108.1	1.49	3.99
Dave Weathers	77	0	1	6	7	26	87.2	87	30	6	40	75	7.70	0.0	1.45	3.08
Aaron Heilman	14	13	2	7	0	0	65.1	79	49	13	41	51	7.03	90.5	1.84	6.75

2.1 한국프로야구에 WHIP의 적용

본 연구의 대상자는 2003시즌 30이닝 이상을 던진 투수 103명을 대상으로 하였다. 그 이유는 규정이닝, 즉 133이닝을 적용시키면 선발투수 위주로 분석이 되어지기 때문이다. 그러므로 중간계투와 마무리를 고려하여 30이닝을 기준으로 하였다. 그리고 투수의 보직은 투수가 등판한 시점을 기준으로 많이 등판한 쪽을 투수의 보직으로 구분하였다. 그러므로 팀에서 구분하는 보직과는 약간

의 차이가 있을 수도 있다. 1982년에서 2002년까지의 팀 자료는 회귀모형에 대한 검정용으로 활용하고자 한다. 그리고 본 연구에서 제시되는 투수들이 허용한 안타 개수, 볼넷 개수 등은 2003년 한국프로야구 정규리그에서 발생한 자료들이다. WHIP 상위 10위 선수의 방어율, 투구이닝, 투수보직은 [표 2]와 같다.

[표 2] WHIP 상위 10위 선수의 방어율, 투구이닝, 투수보직					
선수	팀	방어율	WHIP	이닝	투수보직
이강철	기아	1.98	0.95	68.30	마무리
김현욱	삼성	2.00	0.96	85.60	중간
라이언	삼성	5.02	0.98	83.00	선발
정대현	SK	2.29	0.98	51.00	중간
이원식	기아	3.08	1.08	38.00	중간
오상민	삼성	2.79	1.09	58.00	중간
이상훈	LG	3.34	1.11	56.60	마무리
손혁	두산	3.65	1.12	61.67	선발
조웅천	마무리	1.97	1.15	77.60	마무리
가득염	중간	5.80	1.15	35.60	중간

3. 한국프로야구자료의 통계분석

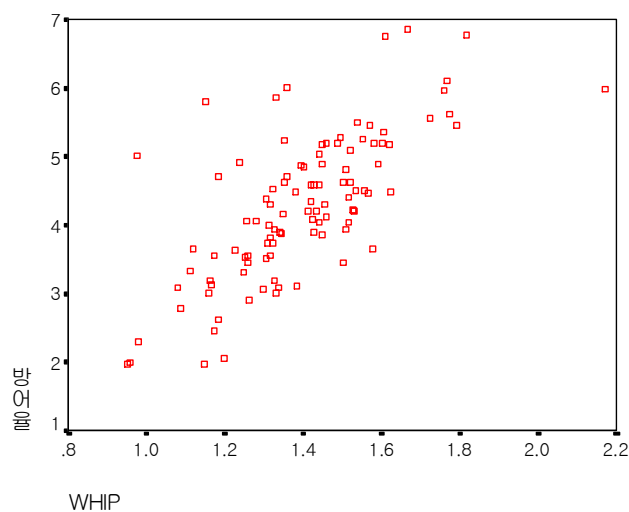
통계패키지 spss10.0(강병서 등, 1998)을 이용하여 2003 한국프로야구자료에서 WHIP와 방어율을 상관분석하고, IW와 IH를 바탕으로 군집분석하여 방어율을 비교분석하고, 투수보직에 따른 WHIP를 비교하였다. 또한 IW와 IH를 독립변수로 방어율을 종속변수로 하는 회귀모형을 추정하여 독립변수를 WHIP로 추정한 회귀모형과 1982년에서 2002년까지 팀 자료를 바탕으로 회귀모형의 적합성을 비교하였다.

3.1 WHIP와 방어율의 상관분석

WHIP와 방어율의 상관계수에 대한 결과와 산점도는 [표 3]과 [그림 1]과 같다. 그 결과에서 보면 상관계수 $r = 0.711$ ($p\text{-value} = .000$)으로 유의수준 1%에서 통계적으로 유의하고, WHIP가 높아질수록 방어율도 높아지는 강한 양의 상관관계가 있었다.

[표 3] WHIP와 방어율의 상관관계			
		승률	WHIP
승률	Pearson 상관계수	1.000	0.711**
	유의확률 (양쪽)	.	.000
WHIP	Pearson 상관계수	0.711**	1.000
	유의확률 (양쪽)	.000	.

** $p < .01$ * $p < .05$



[그림 1] WHIP와 방어율의 산점도

3.2 군집분석을 통한 방어율 비교분석

IW와 IH를 바탕으로 군집분석을 실시한 결과는 [표 4]와 같다. 2군집은 IW와 IH가 모두 낮은 군집이고, 3군집은 2군집과 상반되게 IW와 IH가 모두 높은 군집이며, 1군집은 IW는 낮고, IH가 조금 높은 군집이며, 4군집은 1군집과 상반되는 군집으로 IW는 가장 높고, IH가 조금 낮은 군집이다.

[표 4] 최종군집 유형

	군 집			
	1	2	3	4
IH	.88210	-.71790	1.36321	-.40929
IW	-.57145	-.58835	.99510	1.05211
케이스수	28	38	10	27

2군집과 3군집에 포함되는 선수에 대한 WHIP와 방어율의 자료는 아래 [표 5], [표 6]와 같다.

[표 5] 2군집 소속 선수

선수명	보직	방어율	이닝	WHIP	선수명	보직	방어율	이닝	WHIP
이강철	마무리	1.98	68.3	0.95	정민태	선발	3.31	177	1.25
김현욱	중간	2	85.6	0.96	마정길	중간	4.06	68.6	1.25
라이언	선발	5.02	83	0.98	김진우	선발	3.45	169.3	1.26
정대현	중간	2.29	51	0.98	임창용	선발	3.55	147	1.26
이원식	중간	3.08	38	1.08	이상열	중간	2.91	58.6	1.26
오상민	중간	2.79	58	1.09	키퍼	선발	4.07	77.3	1.28
이상훈	마무리	3.34	56.6	1.11	김승희	중간	4.38	51.33	1.31
손혁	선발	3.65	61.67	1.12	조용준	마무리	3.52	53.6	1.31

선수명	보직	방어율	이닝	WHIP	선수명	보직	방어율	이닝	WHIP
조웅천	마무리	1.97	77.6	1.15	이리키	선발	3.74	159	1.31
가득염	중간	5.8	35.6	1.15	정민철	선발	4	139.6	1.31
존슨	선발	3	69	1.16	최상덕	선발	3.56	126.3	1.31
권준현	마무리	3.19	93	1.16	리오스	선발	3.82	188.6	1.31
송진우	선발	3.12	129.6	1.17	김상진	선발	4.53	89.3	1.32
키피	선발	3.55	91.33	1.17	전준호	선발	3.93	68.6	1.33
전승남	중간	2.45	106.6	1.17	바워스	선발	3.01	143.6	1.33
정성훈	중간	2.62	58.33	1.18	진필중	마무리	3.08	64.3	1.34
에스트라다	선발	4.71	36.3	1.18	안영명	중간	3.89	44	1.34
윤길현	중간	2.06	39.3	1.2	고우석	중간	3.88	58	1.34
신용운	마무리	3.63	119	1.23	김명완	중간	6	42	1.36

[표 6] 3군집 소속 선수

선수명	보직	방어율	이닝	WHIP	선수명	보직	방어율	이닝	WHIP
박명환	선발	5.19	100.67	1.58	서승화	중간	5.57	82.3	1.73
이정훈	중간	5.2	110.6	1.6	마일영	선발	5.97	69.3	1.76
이동학	중간	5.35	77.3	1.6	김성태	중간	6.1	41.3	1.77
이병석	선발	5.18	66	1.62	송은범	중간	6.78	71.6	1.82
최영필	중간	6.86	42	1.67	라형진	중간	5.98	57.6	2.17

이러한 군집 유형에 따른 방어율의 평균비교를 위한 분산분석의 결과를 살펴보면 [표 7]과 같다. 분산분석의 결과를 살펴보면 $p\text{-value}=.000$ 으로 유의수준 1%에서 통계적으로 평균의 유의한 차이가 있는 것으로 나타났고, 사후분석의 결과 2군집, 즉 IW와 IH가 모두 낮은 군집의 평균방어율이 가장 낮았고, 3군집, 즉 IW와 IH가 모두 높은 군집의 평균방어율이 가장 높은 것으로 나타났고, 1군집과 4군집은 통계적으로 평균차이가 없는 것으로 나타났다

[표 7] 군집유형에 따른 방어율 비교 결과

	N	평균	표준편차	F(p-value)	Duncan
1	28	4.7118	.6602	27.791 (.000)**	2 < 4 = 1 < 3
2	38	3.4984	.9450		
3	10	5.8180	.6328		
4	27	4.3811	.7274		
합계	103	4.2849	1.0597		

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

3.3 투수보직에 따른 WHIP의 비교

투수보직에 따른 WHIP의 비교 결과는 [표 8]과 같다. 결과를 살펴보면 $p\text{-value}=.03$ 으로 유의수준 5%에서 통계적으로 평균의 유의한 차이가 있었다. 사후분석결과 선발투수나 중간계투진에 비하여 마무리 투수들의 WHIP가 낮은 것을 알 수 있고, 선발과 중간계투진의 통계적으로 평균차이는 없는 것으로 나타났다.

[표 8] 투수보직에 따른 WHIP 비교 결과

	N	평균	표준편차	F(p-value)	Duncan
선발	41	1.3856	.1618	3.644 (.030)*	마무리 < 선발 = 중간
중간	51	1.4328	.2255		
마무리	11	1.2593	.1547		
합계	103	1.3955	.2007		

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

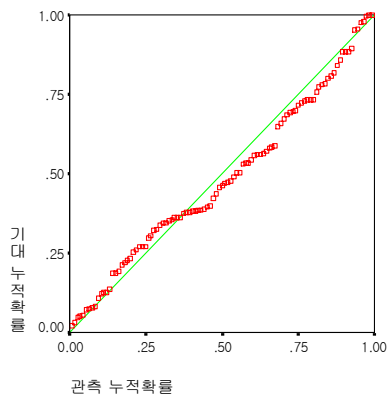
3.4 IW와 IH를 이용한 회귀분석

IH와 IW를 이용하여 방어율을 추정하는 모형을 구축하기 위한 회귀분석의 결과는 [표 9]과 같다. 모형에 대한 검정 결과는 $p\text{-value} = .000$ 으로 유의수준 1%에서 통계적으로 유의하고, $adj-R^2 = .557$ 로 55.7% 정도의 설명력을 가지고 있었다. [그림 2], [그림 3]의 결과로부터 잔차에 대한 정규성, 등분산성에 문제가 없음을 알 수 있고, Dubin-Waston 통계량은 2.027로 독립성에 대한 문제도 없음을 알 수 있다. 추정된 회귀식은 아래와 같다.

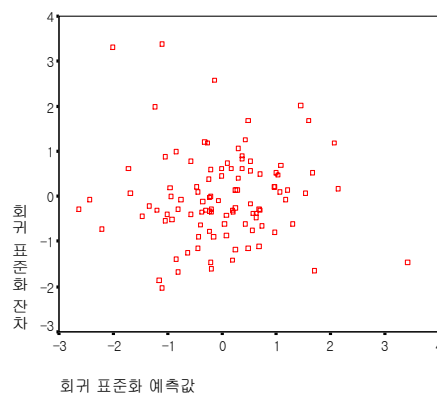
[표 9] IH와 IW를 이용한 회귀모형

모형	비표준화 계수		표준화 계수	t	유의확률
	B	표준오차	베타		
(상수)	-1.750	.535		-3.274	.001
1 IH	5.050	.491	.678	10.287	.000
IW	2.461	.490	.331	5.023	.000
F(p-value)	65.218(.000)**				
R^2	.566				
$adj-R^2$.557				
DW	2.027				

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



[그림 2] 정규확률도



[그림 3] 잔차의 예측치 산점도

$$\text{방어율} = -1.750 + \text{IH} \times 5.050 + \text{IW} \times 2.461$$

IW와 IH의 표준화 계수를 살펴보면 IW가 .331 IH는 .678로 방어율에 미치는 영향이 상대적으로 IW보다는 IH가 높은 것을 알 수 있다.

3.5 WHIP를 이용한 회귀분석

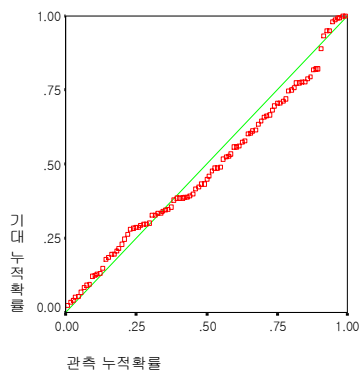
WHIP를 이용하여 방어율을 추정하는 모형을 구축하기 위한 회귀분석의 결과는 [표 10]와 같다. 모형에 대한 검정 결과를 보면 $p\text{-value} = .000$ 으로 유의수준 1%에서 통계적으로 유의하고, $adj-R^2 = .500$ 로 50.0% 정도의 설명력을 가지고 있었다. [그림 4], [그림 5]의 결과로부터 잔차에 대한 정규성, 등분산성에 문제가 없음을 알 수 있으며, *Dubin-Waston* 통계량은 1.910로 독립성에 대한 문제도 없음을 알 수 있다. 따라서 추정된 회귀식은 아래와 같다.

$$\text{방어율} = -.952 + \text{WHIP} \times 3.753$$

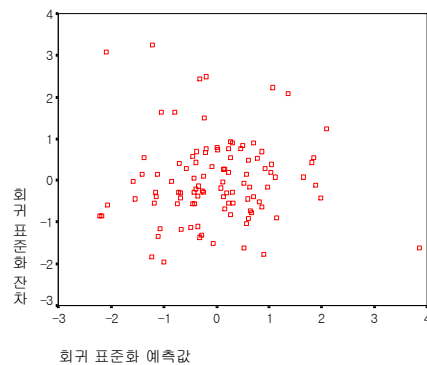
[표 10] WHIP를 이용한 회귀모형

모형	비표준화 계수		표준화 계수		t	유의확률
	B	표준오차	베타			
1	(상수)	-.952	.521		-1.828	.071
	WHIP	3.753	.370	.711	10.155	.000
F(p-value)				103.127(.000)**		
R^2				.505		
$adj-R^2$.500		
DW				1.910		

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



[그림 4] 정규확률도



[그림 5] 잔차의 예측치 산점도

3.6 회귀분석의 모형 검정

회귀모형에 대한 검정과 더불어 독립변수 WHIP를 이용한 회귀분석과 IW와 IH를 이용한 회귀분석을 비교하기 위하여 1982년에서 2002년까지 팀 자료를 활용하기로 한다.

모형확인의 가장 좋은 방법은 새로운 자료를 수집하여 선택된 모형을 검토하는 것이다. 새로운 자료에서 설명변수들의 특정한 값을 모형에 대입했을 때 예측값과 새로운 자료의 관측값과 얼마나 차이가 나는지를 보기 위하여 MSPR(mean squared prediction error)을 사용하며, 그 공식은 다음과 같이 정의된다.

$$\text{MSPR} = \frac{\sum_{i=1}^{n'} (y_i - \hat{y}_i)^2}{n^*},$$

여기서, y_i 는 새로운 자료의 관측값, \hat{y}_i 는 최종모형을 이용한 예측값, n^* 는 자료의 개수.

그러므로 MSPR이 0에 가까울수록 모형이 잘 적합 되었다고 이야기 할 수 있다. 독립변수 WHIP를 이용한 회귀식과 IW와 IH를 이용한 회귀식의 MSPR을 비교하여 보면 IW와 IH를 이용한 MSPR = 0.23이고, WHIP를 이용한 MSPR = 0.36으로 나타났다. 그러므로 두 회귀분석의 설명력의 결과와 마찬가지로 방어율을 예측하기 위한 회귀모형은 독립변수 IW와 IH를 이용한 선형회귀모형이 적합도가 더 좋았다.

4. 결론

“한국프로야구에서 Beane Count 적용에 관한연구”(조영석 등, 2003)를 통하여 승률에 홈런과 볼넷이 미치는 영향을 연구하였다. 홈런과 볼넷은 모두 투수와 타자간의 대결에서 나타난 결과물이다. 야구에서 투수가 차지하는 비중은 팀 전력의 절반이상을 차지한 다고해도 과언이 아닐 정도다.

이런 투수의 능력을 평가하는 자료로 주로 승률이나 투구이닝, 구질, 신체조건, 방어율등을 활용한다. 군집분석 결과에서 보듯이 WHIP를 이루는 요소인 IW와 IH가 낮은 군집이 IW와 IH가 높은 군집보다 방어율이 더 낮은 것으로 나타났으며, 투수보직에 대한 WHIP는 마무리투수가 가장 낮은 것으로 나타났고, 회귀분석의 결과 IH가 IW보다는 방어율에 더 영향을 주고 있고, 독립변수 IW와 IH가 WHIP를 이용하는 것보다 모형의 설명력이 높은 것으로 나타났다. 최근에 메이저리그에서는 WHIP를 중요한 수치로 사용하고 있고, WHIP는 방어율을 간접적으로 측정 할 수 있을 뿐 아니라 WHIP를 이루는 요소인 IW와 IH로 투수의 능력을 측정할 수 있다 앞으로 투수의 능력을 측정할 수 다른 변수의 연구가 필요하며, 한국프로야구에서 매 시즌 분석을 통하여 WHIP에 대한 연구를 하고자 한다.

참고문헌

- [1] 조영석, 조용주 (2003). 한국프로야구에서 Beane Count 적용에 관한 연구, *Journal of the Korean Data Analysis Society*, Vol. 5, No. 3, pp. 649-658.
- [2] 강병서, 김계수 (1998). *사회과학 통계분석, SPSS 아카데미*
- [3] 사단법인 한국야구위원회 (2003). *프로야구 기록 대백과*, 사단법인 한국야구위원회.
- [4] Rov Never (2002). <http://www.espn.com>.
- [5] 스포츠 서울 전문기자 칼럼 (2003). <http://www.sportsseoul.com>

[2004년 8월 접수, 2004년 10월 채택]

K C I

Study about the influence that WHIP has on ERA
in 2003 season Korean professional baseball

*Young-Suk Cho*¹, *Yong-Ju Cho*²

Abstract

Baseball is very important pitcher in game. The performance of pitcher is conjectured win percentage(Win%), wins, innings pitched(IP), strikeout per innings, holds(HLD), Saves(S), quality start(QS) and earned runs average(ERA). Especially, ERA is very important for performance of pitcher. We used this inning a walk(IW) with this inning a hit(IH) in study and executed cluster analysis, and compared ERA(earned run average) according to cluster. Also, we compared WHIP according to part of a pitcher.

An important element to achieve WHIP(walk+hit inning pitch) is IW and IH. Those were utilized for regression analysis. Also we compared WHIP utilized for regression analysis.

Keywords : ERA, WHIP, Regression Analysis, Cluster Analysis.

¹Assistant professor, School of Free Major, Miryang National University 1025-1, Naeidong, Miryang, Kyungnam 627-702, Korea. E-mail : choys@mnu.ac.kr

²Department of Statistics Graduate School, Yeungnam University Gyeongsan 712-749, Korea. E-mail : nuriya@empal.com