

2017년 한국프로야구 타자력 예측모형 개발

홍중선¹ · 신동식²

¹²성균관대학교 통계학과

접수 2017년 4월 3일, 수정 2017년 5월 8일, 게재확정 2017년 5월 15일

Abstract

2017년 한국프로야구 (KBO) 시즌이 개막되어 2017년도 선수들의 경기력 중에서 타자력 예측모형을 설정하고 분석한다. KBO 투수자료의 빈약함으로 규정타석을 만족한 미국프로야구 (MLB)와 KBO 타자 선수들을 대상으로 선정된 여덟 개의 타자력 지표와 수집된 과거 3개년도 자료를 바탕으로 2016년 타자력 예측모형을 개발한다. 예측모형이 MLB와 KBO 모두에 적합하고 특히 MLB에 비교하여 KBO 환경에 더욱 적합함을 발견한다. 타당성이 만족한 예측모형에 대하여 MLB와 KBO 타자력을 비교 분석하고, 2017년 타자력을 추정할 수 있다. 그리고 MLB와 KBO 타자력 예측모형과 타자 선수들의 연령과의 분석으로부터 대부분의 유의한 타자력 지표들은 연령과 관계가 없다고 판단된다.

주요용어: 가중평균, 종단연구, 지표, 투수력, 타자력.

1. 서론

2017년 한국프로야구 (KBO) 시즌이 작년에 이어 총 10개의 프로야구팀 (KT, 롯데, 넥센, LG, KIA, 두산, NC, 한화, SK, 삼성)으로 3월 31일에 개막되었다. 2016년도 시즌에 FA 100억 시대와 함께 한국 프로스포츠 사상 최초로 800만 관중을 돌파하면서 2017년도 시즌은 열기를 더해갈 전망이다. 여기에 야구팬들의 관심사에는 ‘올해 우승팀을 예측’하거나 ‘자신이 응원하는 팀의 성적’이며, ‘2017년 타자 최형우와 이대호의 타자력 지표 중에서 장타율은 작년보다 좋아질까? 또는 투수 니퍼트와 차우찬의 투수력 지표 중에서 방어율은 어떻게 예측될까?’라는 식의 어떤 특정한 선수의 기량을 추정하여 선수가 소속된 팀에 얼마나 기여할지를 생각해 보는 것이 중요한 관심사가 된다. 따라서 본 연구에서는 2017년도 시즌 선수들의 경기력 예측모형을 설정하고 분석하고자 한다. 2017년 예측하기 위하여 우선 과거 자료인 2015년까지의 자료를 바탕으로 2016년 경기력을 예측하고, 예측자료와 결과를 비교하여 모형의 타당성을 검토하고 이를 바탕으로 2017년 예측모형을 개발하고자 한다.

야구 선수들의 경기력에 관한 기존 연구문헌으로 Lee (2014a, 2014b)는 2000년부터 2013년까지의 자료의 평균을 바탕으로 주성분분석을 이용하여 새로운 투수지표와 타자지표 제안하였고, Hong 등 (2016)은 2015년도 자료를 이용하여 새로운 타자지표 제안 및 연봉과의 관계분석을 실시하였다. Lee (2007)는 1996부터 2006년까지 자료의 평균으로 분석하였고, Kim 등 (2013)은 각 타석에서의 안타 여부가 이전 타석에서의 결과와 관련이 있는지를 살펴보는 연구를 2011년도 자료를 이용하였다. Cho와 Lee (2015)는 2013년도 자료를 이용하여 타자력 지표를 순수한 타격능력을 모수로 간주하여 베이지안 기법으로 추정하였다. 선수들의 경기력과 연봉과의 관계를 분석하는 문헌 중에서 Kim (2001,

¹ 교신저자: (03063) 서울. 종로구 성균관로 25-2, 성균관대학교 통계학과, 교수. E-mail: cshong@skku.edu

² (03063) 서울. 종로구 성균관로 25-2, 성균관대학교 통계학과, 대학원생.

2002)은 3개년도 각각의 자료를 사용하고, Park (2008)은 2006년도 자료를 이용하였으며, Seung과 Kang (2012)은 2009년, 2010년도 자료를 각각 사용하면서 분석하였다. 타자 경기력과 득점과의 관계에 대한 연구 문헌 중에서 Kim (2012)은 2007년도부터 2011년도 자료를 독립적으로 각각 사용하였고, Lee (2014c)는 1982년도부터 2013년도까지의 자료를 독립적으로 각각 사용하고, Kim과 Kim (2014)은 1982년도부터 2013년까지의 자료를 독립적으로 각각 사용하였다. 이와 같이 다음 시즌의 경기력을 예측하는 대부분 연구문헌들은 직전연도의 자료만을 이용하거나 2년 이상의 과거 자료를 이용한다 하더라도 과거 자료들의 평균을 이용하여 분석하였다. 본 연구에서는 다른 문헌연구와는 다르게 직전연도까지의 과거 자료 모두를 이용하여 종단적인 연구 (longitudinal study)를 진행하고자 한다. 우선 우리나라의 자료보다 방대한 미국프로야구 (MLB)와 한국프로야구 (KBO)의 홈페이지를 통하여 공개된 자료 중에서도 가능한 오랜 기간 축적된 자료를 수집하여 경기력 예측 모형을 설정하고자 한다.

본 연구에서의 분석 대상인 선수는 시즌 별 규정타석 (plate appearances; PA, 경기 수*3.1)을 만족한 타자와 규정이닝 (inning pitched; IP, 경기 수*1.0)을 만족한 투수이다. 2011년부터 2016년까지 활동한 타자와 투수는 Table 1.1과 같다.

Table 1.1 Number of batters and pitchers satisfying PA/IP (qualified)

	KBO						MLB					
	2011	2012	2013	2014	2015	2016	2011	2012	2013	2014	2015	2016
Batter	37	40	48	55	51	55	145	143	140	146	141	146
Pitcher	16	22	25	23	20	17	94	88	81	84	78	73

과거 자료를 이용하여 2016년 경기력을 예측하고 2016년 선수의 경기력과의 차이를 구하여 분석하고자 하는데, 2011년부터 2016년까지 지속적으로 활동한 선수들은 Table 1.2와 같다.

Table 1.2 Number of batters and pitchers who run till 2016

Period	Batter		Pitcher	
	KBO	MLB	KBO	MLB
2011 ~ 2016	9	21	0	10
2012 ~ 2016	13	31	0	14
2013 ~ 2016	17	45	2	19

Table 1.2를 살펴보면, MLB 투수 자료의 표본 크기는 적절하나, KBO 투수의 경우 선수층이 얇아 2016년도 시즌부터 2013년도 시즌까지 4개년도 자료를 사용할 때는 2명, 5개년도 이상의 자료를 사용할 때는 0명으로 연구를 진행하기 위한 자료수가 절대적으로 부족하다. 따라서 본 연구에서는 투수의 경기력 예측모형 개발은 제외한다. 그리고 2015년까지 MLB와 KBO에서 활동한 타자 자료의 표본크기는 충분하다고 간주하여 MLB와 KBO 타자에 대한 타자력 지표의 예측모형을 설정하고 분석하고자 한다.

본 연구의 구성은 다음과 같다. 2절에서는 MLB와 KBO 홈페이지를 통하여 수집된 자료의 성격을 바탕으로 타자력 예측모형을 설정한다. 그리고 예측모형에 포함된 타자력 지표들을 선정한다. 우선 방대한 자료를 갖고 있는 MLB 자료를 대상으로 설정한 예측모형의 적합성과 타당성을 검증한 다음에 이 모형을 KBO 자료에 적용하여 한국프로야구 타자의 타자력 예측모형을 설정하고 토론한다. 3절에서는 타자력 예측하는데 중요한 선수 인자 중의 하나인 선수의 연령을 고려하여 타자력 예측모형과 선수의 연

령과의 관계를 분석하고 토론한다. 그리고 4절에서는 타자력 예측모형에 대하여 종합적으로 분석한 결과를 설명하고 2017년 타자의 경기력 예측에 대하여 결론을 유도한다.

2. 타자력 예측모형

타자의 경기력 예측모형을 개발하기 위하여 Table 1.2를 살펴보면, 2011년 자료부터 2015년까지 KBO에서 활동한 선수가 9명이며, 2012년부터는 13명, 그리고 2013년부터도 17명으로 회귀모형 또는 시계열 모형을 적용하기에는 작은 표본이므로 적절하지 않다고 판단된다. 따라서 2013년부터 2015년까지 KBO에서 활동한 17명의 타자 자료인 타자력 지표 (Batting Index)에 대하여 가중평균을 이용한 2016년 타자력 지표의 예측모형 (\widehat{Index}_t)을 다음과 같이 설정한다.

$$\widehat{Index}_t = w_{t-1}Index_{t-1} + w_{t-2}Index_{t-2} + w_{t-3}Index_{t-3}, \quad (2.1)$$

여기서 시점인 t 는 2016년이며 $\sum_{i=t-3}^{t-1} w_i = 1$, 각 $w_i \geq 0$ 이다. 본 연구에서는 야구의 승패 핵심이라고 볼 수 있는 출루와 득점에 관계된 다음의 여덟 개 타자력 지표에 대하여 고려하며, 출루와 득점에 직접적인 관계없는 지표와 여덟 개의 타자력 지표의 계산식에서 중복 사용하는 지표는 제외하고 예측모형을 개발한다.

OBP (On Base Percentage; 출루율): 타수에 대한 출루의 비율로써 타자가 아웃되지 않고 베이스 출루율인 ((안타+사사구)/(타수+사사구+희생플라이))를 백분율로 나타낸 지표이다.

AVG (Average; 타율): 타수에 대한 안타의 비율 (안타/타수)로써 타격 정확도를 의미한다.

SLG (Slugging Percentage; 장타율): 단타를 1, 2루타를 2, 3루타를 3, 홈런을 4로 적용하여 타수로 나누어 ((단타+2*2루타+3*3루타+4*홈런)/타수)로써 진루의 기대값으로 해석한다.

ISO (Isolated Power; 순장타율): 장타율에서 타율을 뺀 값 (장타율-타율)으로 순수한 장타에 대한 지표이다.

BABIP (Batting Average on Balls In Play; 인플레이 타구의 타율): 인플레이로 이어진 타구에 대한 타율로 ((총안타수-홈런)/(타수-삼진-홈런+희생플라이))이다.

BB/K (Base on Balls per Strike Out; 볼넷/삼진 비율): 타자가 얻은 볼넷의 수를 삼진 수로 나눈 값 (볼넷/삼진)으로 타자의 출루능력을 평가하는 지표이다.

RC27 (Runs Created 27; 한 경기당 득점생산): 특정 선수가 한 경기 모든 타석에 설 경우 팀이 몇 점을 득점할 것인가를 평가하는 지표이다.

$$RC / (\text{타석} - \text{안타} + \text{희생타} + \text{희생플라이} + \text{도루실패} + \text{병살타}) * 27,$$

여기서 $RC = ((2.4C + A)(3C + B)) / 9C - 0.9C$, $A = \text{안타} + \text{볼넷} - \text{도루실패} + \text{사구} - \text{병살타}$, $B = 1.125(1\text{루타}) + 1.69(2\text{루타}) + 3.02(3\text{루타}) + 3.73\text{홈런} + 0.29(\text{볼넷} - \text{고의볼넷} + \text{사구}) + 0.492(\text{희생타} + \text{희생플라이} + \text{도루}) - 0.04\text{삼진}$, $C = \text{타수} + \text{볼넷} + \text{사구} + \text{희생타} + \text{희생플라이}$.

WAR (Wins Above Replacement; 대체선수 대비 승리 기여도): 대체선수에 비해 얼마나 많은 승리에 기여했는가를 나타내는 지표로 선수의 모든 요소를 계산해서 종합한 통계량이며 MLB의 환경에 맞추어서 구성되어있다. 수치가 높을수록 영향력 있는 선수로 해석한다.

$$(\text{타격기여도} + \text{수비기여도} + \text{포지션 조정} + \text{주루기여도}) / \text{승리시득점}.$$

본 연구에서 제안한 예측모형을 설정하기 위하여 가중값에 대한 제약조건을 해제하고 다양한 조건 하에서도 진행하였고 3년 기간의 가중평균뿐만 아니라 4년과 5년 기간의 가중평균도 고려해보았으며 2000년부터 2016년까지의 예측모형도 모두 고려해보았으나, 식 (2.1)의 예측모형이 가장 적합하였기 때문에 2016년도로 타자력 예측모형을 제안한다.

2016년 타자력 예측모형 (2.1)식을 구성하는 각각의 가중값 w_{13} , w_{14} , w_{15} 의 단위를 0.1 간격으로 설정하였으며, 합이 1이 되는 값으로써 실제 2016년 자료인 $Index_{16}$ 과의 차이를 가장 작게 만드는 최적의 가중값을 산출하고, 이를 이용한 예측모형과의 차이 $Diff = Index_{16} - \widehat{Index}_{16}$ 를 이용하여 두 표본 비교분석 (t -test)을 실시한다. 본 연구에서 사용하는 여덟 가지 지표는 예측모형을 구성하는 가중값의 변별력을 높이기 위하여 연도별로 표준화시키지 않은 절대적인 성적을 이용한다.

우선 MLB 타자의 타자력 예측모형 분석을 실시하면서 타당성 분석을 한 후에 KBO 타자에 적용하고자 한다.

2.1. MLB 타자력 예측 분석

2013년부터 2015년까지 MLB에서 활동한 45명에 대한 여덟 종류의 타자력 지표에 대하여 2016년 타자력 지표의 예측모형 (2.1)의 결과는 Table 2.1과 같다.

Table 2.1 Prediction analysis of MLB batting index

	w_{13}	w_{14}	w_{15}	t	$p - value$
OBP	1	0	0	0.7224	0.4739
AVG	0.9	0.1	0	0.0241	0.9809
BABIP	0.3	0.2	0.5	0.0043	0.9966
ISO	0	0	1	2.5399	0.0147
SLG	0	0	1	2.2839	0.0273
BB/K	1	0	0	-0.0072	0.9943
RC27	1	0	0	1.1852	0.2423
WAR	0	0	1	-0.0514	0.9592

타자력 예측모형을 바탕으로 “2016년 예측값과 2016년 타자력지표 자료의 차이는 없다”라는 귀무가설에 대한 두 표본 비교분석 결과에서 유의수준 $\alpha = 0.05$ 에서 MLB의 타자력 지표 중에서 ISO와 SLG는 귀무가설을 기각하여 예측모형이 적절하지 않으므로 MLB의 타자력 예측모형 분석에서 제외하고, 나머지 여섯 개의 지표인 OBP, AVG, BABIP, BB/K, RC27, 그리고 WAR에 대하여 분석한다. 그 중 네 개의 지표를 Figure 2.1에 구현하였다.

Figure 2.1에서 실선은 2013년도부터 2016년까지의 실제 평균값이며 2016년에 *로 표시한 점은 예측모형을 바탕으로 산출된 예측값을 나타낸 것이다. Table 2.1의 결과와 마찬가지로 Figure 2.1에서도 최적의 가중값을 통해 산출한 \widehat{Index}_{16} 과 실제 결과인 $Index_{16}$ 과의 차이($Diff$)를 통한 예측력을 확인할 수 있는데 예측모형으로 적합한 지표로 판단한 MLB의 OBP, AVG, BABIP, BB/K, RC27, 그리고 WAR은 \widehat{Index}_{16} 과 $Index_{16}$ 의 간격이 매우 가까운 것으로 확인할 수 있는 반면 예측모형으로 적합하지 않다고 판단한 MLB의 ISO, SLG는 \widehat{Index}_{16} 과 $Index_{16}$ 의 간격이 큰 것으로 확인할 수 있다. 또한 설정한 예측모형 (2.1)이 적절하다고 판단되는 타자력 지표 중에서 WAR 이외의 지표들은 2015년인 직전연도 자료보다는 2014년과 2013년도의 자료에 가중값이 집중되어 있어 두 개년도의 자료에 민감하게 반응하고 있으며 자세히 살펴보면, OBP, AVG, BABIP, BB/K, 그리고 RC27 지표는 2013년 자료에

더욱 민감하게 반응하여 2016년 예측을 2014년 보다는 2013년도 자료로 예측하는 것이 바람직하다고 판단된다. 그리고 야구선수의 모든 요소를 계산해서 종합한 통계량이며 MLB의 환경에 적절한 WAR 지표만은 직전연도인 2015년 자료에 가장값이 집중되어 있어 2015년 자료에 민감하게 반응함을 파악할 수 있다.

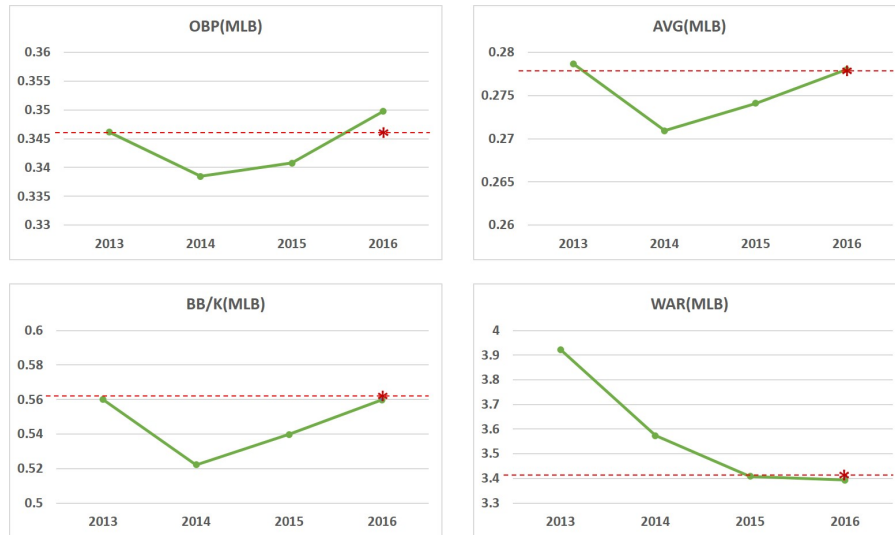


Figure 2.1 MLB batting index

그러므로 본 연구에서 제안한 (2.1)모형은 MLB 타자력 예측하는데 충분히 적합한 모형이며 타자력 예측모형으로 타당하다고 결론내릴 수 있다. 다음으로는 MLB 타자력 예측모형 (2.1)을 KBO 타자에 적용하여 타자력 예측하고자 한다.

2.2. KBO 타자력 예측 분석

2013년부터 2015년까지 KBO에서 활동한 17명에 대한 여덟 종류의 타자력 지표에 대하여 2016년 타자력 지표의 예측모형의 결과는 Table 2.2와 같다.

Table 2.2 Prediction analysis of KBO batting index

	w_{13}	w_{14}	w_{15}	t	$p - value$
OBP	0	1	0	1.3947	0.1822
AVG	0	1	0	1.0009	0.3318
BABIP	0	1	0	0.4364	0.6684
ISO	0	1	0	2E-16	1.0000
SLG	0	1	0	0.4240	0.6772
BB/K	0	1	0	1.5734	0.1352
RC27	0	1	0	1.2556	0.2273
WAR	0	0	1	1.8027	0.0903

2016년 예측모형으로부터 추정한 예측값과 2016년 타자력지표 자료의 차이는 없다라는 귀무가설에 대한 두 표본 비교분석 결과에서 유의수준 $\alpha = 0.05$ 에서 KBO의 모든 타자력 지표는 귀무가설을 채택하여 차이가 없으며 따라서 예측모형이 충분히 적절하다고 해석할 수 있으므로 KBO의 타자력 예측모형 분석에서 모든 여덟 개의 지표인 OBP, AVG, BABAIP, ISO, SLG, BB/K, RC27, 그리고 WAR에 대하여 분석한다. 그 중 네 개의 지표를 Figure 2.2에 표현하였다.

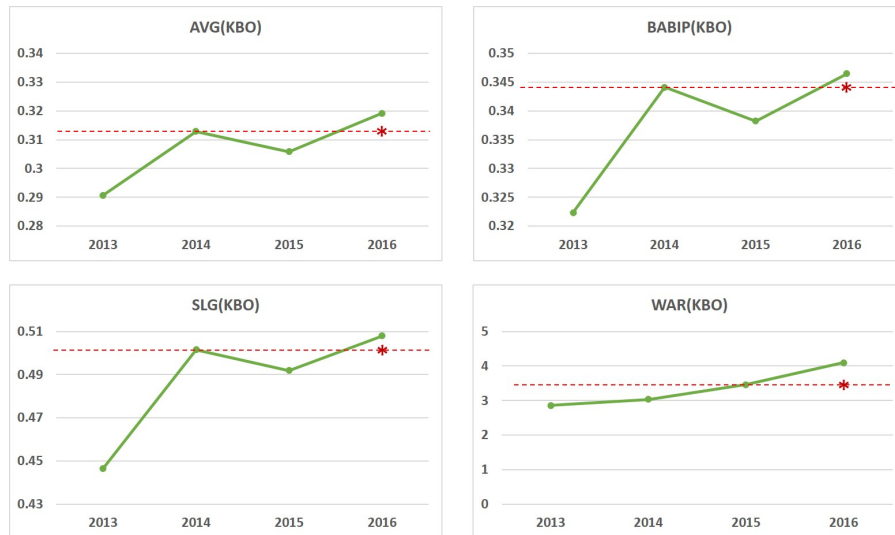


Figure 2.2 KBO batting index

Figure 2.2에서 실선과 *로 표시한 점은 각각 2013년도부터 2016년까지의 실제 평균값과 2016년 예측값을 나타낸다. Figure 2.1를 바탕으로 예측모형으로 적합한 지표로 판단되는 KBO의 OBP, AVG, BABAIP, ISO, SLG, BB/K, RC27, 그리고 WAR은 \widehat{Index}_{16} 과 $Index_{16}$ 의 간격이 매우 가까운 것으로 확인할 수 있다. 그리고 Table 2.2와 Figure 2.2를 살펴보면, 설정한 예측모형 (2.1)식이 적절하다고 판단되는 타자력 지표 중에서 WAR 이외 OBP, AVG, BABAIP, ISO, SLG, BB/K, 그리고 RC27 지표는 유의수준 $\alpha = 0.05$ 에서 모두 2014년 자료에 1이라는 큰 가중값이 집중되어 있어 매우 민감하게 반응하고 있으며 이것은 2016년 예측하기 위해서는 직전연도인 2015년과 2013년 자료보다는 2014년의 자료만으로 충분히 예측 가능하다고 판단할 수 있다. 그리고 KBO 환경에 적용한 WAR 지표만은 MLB에서와 마찬가지로 직전연도인 2015년 자료에 민감하게 반응함을 파악할 수 있다.

MLB와 KBO의 예측모형에 대하여 다음과 같은 세 가지 차이점이 존재한다. 첫째, MLB의 예측모형과 비교하여 KBO의 예측모형에서는 모든 타자력 지표들이 유의한 결과를 갖고 있다. 즉 MLB 예측모형에서는 여덟 개 중에서 6개 지표인 OBP, AVG, BABIP, BB/K, RC27, 그리고 WAR 지표들이 예측모형에 유의한 지표이며, KBO의 예측모형에서는 모든 여덟 개의 지표인 OBP, AVG, BABAIP, ISO, SLG, BB/K, RC27, 그리고 WAR 지표들이 모두 예측모형에 유의한 지표이다. 두 번째로는, KBO의 예측모형에서는 7개의 지표 (OBP, AVG, BABAIP, ISO, SLG, BB/K, RC27)가 2014년도의 자료에 가중값이 큰 값으로 집중되어 있어 2014년 자료에 매우 민감하게 반응하고, MLB의 예측모형에서는 5개의 지표 (OBP, AVG, BABAIP, BB/K, RC27)가 2013년도의 자료에 민감하게 반응하는 것을 탐색할

수 있다. 세 번째로는 야구선수의 모든 요소를 계산해서 종합한 통계량이며 MLB와 KBO 모두에 적절한 WAR 지표는 직전연도인 2015년 자료에 가중값이 집중되어 있으며 이것은 직전연도인 2015년 자료에 민감하게 반응하기 때문에 2016년 예측하기 위해서는 직전연도인 2015년 자료가 충분하다고 판단한다. 그러므로 본 연구를 통해 설정한 예측모형 (2.1)이 MLB에 비하여 KBO 환경에 더욱 적합한 모형이라고 판단할 수 있다. 그리고 전반적으로 2016년 예측한 타자력 지표들은 예측값보다 약간 큰 값으로 추정할 수 있다.

3. 연령과의 관계분석

야구 예측 시스템에서 빼놓을 수 없는 중요한 개념인 James (1987)의 노화곡선 (aging curve)라고 일컫는 선수들의 연령과의 관계를 고려하여 선수들의 연령과 경기력 특히, 타자력과의 관계를 살펴본다. 본 연구의 예측모형 (2.1)을 통해서 얻은 차이 $Diff = Index_{16} - \widehat{Index_{16}}$ 와 2장에서 분석한 MLB와 KBO 타자선수들의 연령에 대하여 연구한다.

3.1. MLB 타자 연령과의 관계

MLB에서 예측모형을 만족한 여섯 개의 지표 (OBP, AVG, BABIP, BB/K, RC27, WAR) 중에서 대표적으로 네 개의 지표 (OBP, AVG, RC27, WAR)들에 대한 차이 $Diff = Index_{16} - \widehat{Index_{16}}$ 와 MLB 타자선수들의 연령과의 관계를 Figure 3.1과 같이 표현하였다.

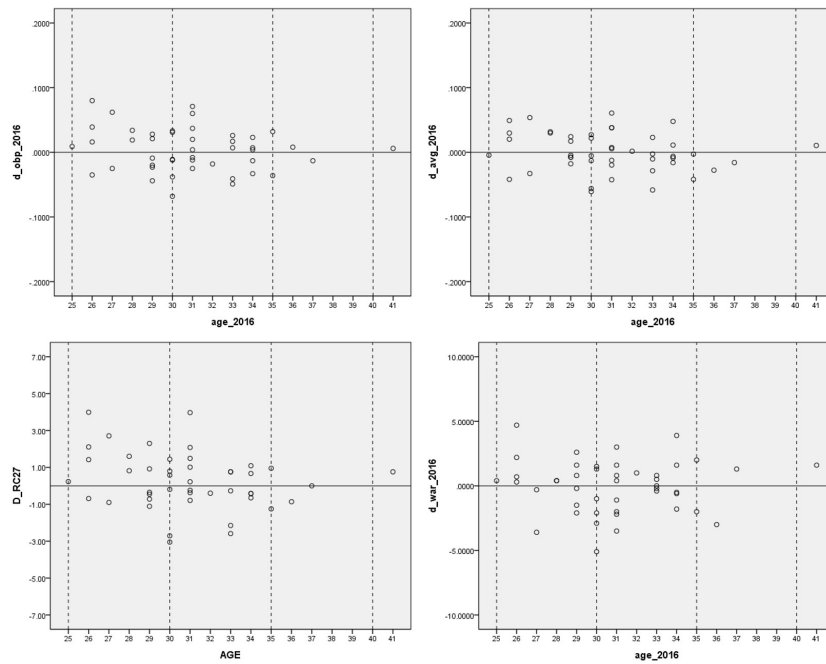


Figure 3.1 MLB batting index and age

Figure 3.1에서의 네 개의 지표뿐만 아니라 여섯 개의 지표 (OBP, AVG, BABIP, BB/K, RC27, WAR) 모두 연령이 증가함에 따라 예측값과의 차이 $Diff$ 가 증가하거나 감소하는 형태를 발견할 수 없다. 즉 연령이 증가해도 차이는 0을 중심으로 특정한 경향 없이 무작위로 퍼져있음을 탐색할 수 있다. 따라서 MLB에서의 예측모형 (2.1)은 선수들의 연령과 관계가 없다고 판단된다.

3.2. KBO 타자 연령과의 관계

KBO에서 예측모형을 만족한 여덟 개의 지표 (OBP, AVG, BABIP, ISO, SLG, BB/K, RC27, WAR) 중에서 대표적으로 네 개의 지표 (OBP, AVG, BABIP, RC27)들에 대한 차이 $Diff = Index_{16} - \widehat{Index}_{16}$ 와 MLB 타자선수들의 연령과의 관계를 Figure 3.2에 표현하였다.

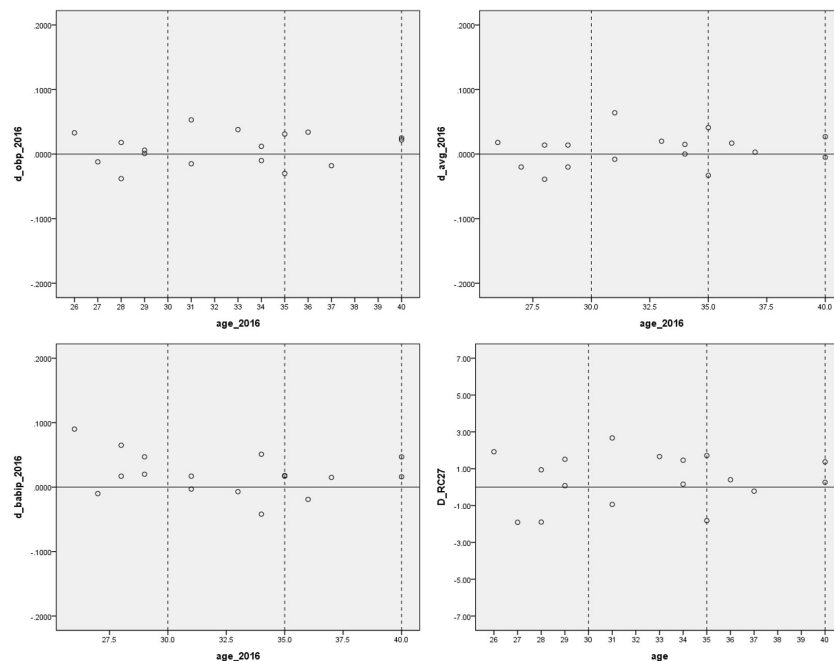


Figure 3.2 KBO batting index and age

Figure 3.2를 통해서 네 개의 지표뿐만 아니라 여덟 개의 지표 모두 연령이 증가함에 따라 예측값과의 차이 $Diff$ 는 0을 중심으로 편이 없이 무작위로 퍼져있으므로 연령이 증가해도 차이의 경향이 존재하지 않는다고 판단할 수 있다. 따라서 KBO의 예측모형 (2.1)은 MLB에서의 예측모형과 유사하게 선수들의 연령과 관계가 없다고 판단된다. 그리고 Figure 3.1과 Figure 3.2를 비교하면, 35세 이상의 고령 타자들의 경우 KBO 선수들은 6명으로 $6/17=0.35$, MLB 선수들은 5명으로 $5/40=0.11$ 로 고령타자들은 KBO 선수들이 MLB의 선수들보다 상대적으로 두텁게 형성되어 있는 것을 확인할 수 있다. 이는 평균적으로 27~30세에 절정을 맞이하고 감소하는 추세를 겪는 Aging Curve 이론을 KBO환경에서는 크게 적용되지 않음을 확인할 수 있다. 그러므로 KBO 타자들은 은퇴시기에 가까워질수록 떨어지기 마련인 경기력을 잘 관리하기 때문에 경기력 감소속도를 낮춰 계속하여 규정타석에 이름을 올릴 수 있는 것이라

고 판단할 수 있다.

4. 결론

본 연구는 규정타석, 규정이닝을 만족한 MLB와 KBO 선수들을 대상으로 수집 가능한 과거년도 자료를 바탕으로 경기력 지표들의 예측모형을 개발하였다. MLB 투수의 자료는 충분하지만 KBO 투수 중에서 최소한 3개년 이상 규정이닝을 만족한 투수는 두 명밖에 되지 않으므로 투수력 예측은 어렵기 때문에 생략하고, 타자력 예측모형에 대하여 연구하였다. 우선 MLB와 KBO 자료 중에서 야구 승패 핵심이라고 볼 수 있는 출루와 득점에 관계되는 여덟 개의 타자력 지표를 선정하였다. 그리고 KBO 자료 중에서 2013년도부터 2015년의 3개년 자료를 바탕으로 (2.1)과 같은 예측모형을 개발하였다. 결과를 분석하면서 MLB와 KBO 타자력을 비교하였다. 그리고 타자력 예측모형과 타자 선수들의 연령과의 관계를 설명하였다.

MLB 예측모형에서 OBP, AVG, BABIP, BB/K, 그리고 RC27 타자력 지표는 2015년인 직전년도 자료보다는 2014년과 2013년도의 자료에 민감하게 반응하고 있으며 특히 2013년 자료에 더욱 민감하게 반응하여 2016년 예측을 직전연도인 2015년도 보다는 2014년과 2013년 자료로 예측하는 것이 적절하며, 그 중에서 2014년 보다는 2013년도 자료로 예측하는 것이 바람직하다고 판단된다. 그리고 WAR 지표만은 직전연도인 2015년 자료로 예측하는 것이 충분하다고 판단할 수 있다.

KBO의 예측모형을 구성하는 여덟 개의 타자력 지표 모두가 유의한 지표임을 발견하였으며, 그 중에서 OBP, AVG, BABIP, ISO, SLG, BB/K, 그리고 RC27 지표는 2016년 예측하기 위해서는 2014년의 자료만으로 충분히 예측 가능하다고 판단할 수 있으며, WAR 지표만은 MLB에서와 마찬가지로 직전연도인 2015년 자료로 예측할 수 있음을 파악할 수 있다.

본 연구를 통해 설정한 타자력 예측모형 (2.1)이 MLB와 KBO 모두에 적합한 모형임을 발견하였고, 특히 MLB에 비교하여 KBO 환경에 더욱 적합한 모형이라고 판단할 수 있다. 그리고 전반적으로 2016년 예측한 타자력 지표들은 예측값보다 약간 큰 값으로 추정할 수 있다. 그러므로 본 연구에서 제안한 KBO 타자력 예측모형을 바탕으로 2017년 타자력을 예측할 수 있다.

KBO의 2016년 타자력 예측모형에서 가장 값은 2014년도 자료에 집중되어 있음을 발견할 수 있는데 이것은 2016년 KBO 환경이 2014년과 유사하기 때문이며 2014년도와 2016년 공통적인 현상을 ‘투저타고’ 현상으로 해석할 수 있다. 2017년 2월에 진행되었던 WBC (World Baseball Classic) 경기에서 한국팀이 1라운드에서 탈락하는 충격적인 상황이 발생하였는데 원인 중 하나로는 외국으로 진출한 류현진이나 김광현 같은 특급 투수들의 부재로 저조한 KBO 투수들의 현주소를 꼽는다. 대형 투수뿐만 아니라 대형 타자들을 또한 KBO를 떠나 해외로 진출해 있는 시점에도 불구하고 KBO의 2017년 상황은 상대적으로 타자력이 높은 경기를 2016년에 해왔기 때문에 2017년에도 여전히 ‘투저타고’ 현상이 계속 유지될 것이라고 관망하면서 2017년 타자력 예측은 2016년 경기력보다 약간 높은 값으로 증가 추세로 예측할 수 있다.

MLB 예측모형에서의 여섯 개의 지표 (OBP, AVG, BABIP, BB/K, RC27, WAR) 그리고 KBO 모형에서의 여덟 개의 지표 (OBP, AVG, BABIP, ISO, SLG, BB/K, RC27, WAR) 모두는 연령이 증가해도 차이는 0을 중심으로 무작위로 퍼져있음을 탐색할 수 있으므로 KBO의 예측모형은 MLB에서의 예측모형과 유사하게 선수들의 연령과 관계가 없다고 판단된다. 그리고 35세 이상의 타자들의 경우 KBO 선수들이 MLB의 선수들보다 상대적으로 두텁게 형성되어 있다. 이는 KBO 선수들이 MLB 투수들보다도 경기력을 경기력 이외의 외적 요인도 잘 관리하는 것으로 판단할 수 있다.

References

- Cho, Y. J. and Lee, K. H. (2015). Bayesian estimation of the Korea professional baseball players' hitting ability based on the batting average. *Journal of the Korean Data & Information Science Society*, **26**, 197-207.
- Hong, C. S., Kim, J. Y. and Shin, D. S. (2016). Alternative hitting ability index for KBO. *Journal of the Korean Data & Information Science Society*, **27**, 677-687.
- James, B. (1987). *The Bill James Baseball Abstract 1987*, Ballantine Books, New York.
- KBO. <http://www.koreabaseball.com/Record/Player/HitterBasic/Basic01d.aspx>.
- Kim, B., Park, Y. and Jang, N. (2013). Study for independence of hits in professional baseball games. *Journal of the Korean Data & Information Science Society*, **24**, 1421-1428.
- Kim, E. S. (2001). The relation of game performance and annual salary for Korean professional baseball players. *Journal of Korean Sociology of Sport*, **14**, 15-24.
- Kim, E. S. (2002). The relationship of game performance and annual salary for Korean professional baseball pitchers. *Journal of Korean Sociology of Sport*, **15**, 95-104.
- 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.
- Kim, H. J. and Kim, Y. H. (2014). Explanation of run productivity using weighted adjusted ops in Korean professional baseball. *Korean Journal of Applied Statistics*, **27**, 731-741.
- Lee, J. T. (2014a). Pitching grade index in Korean pro-baseball. *Journal of the Korean Data & Information Science Society*, **25**, 485-492.
- Lee, J. T. (2014b). Measurements for hitting ability in the Korean pro-baseball. *Journal of the Korean Data & Information Science Society*, **25**, 349-356.
- Lee, J. T. (2014c). Estimation of OBP coefficient in Korean professional baseball. *Journal of the Korean Data & Information Science Society*, **25**, 357-363.
- Lee, Y. H. (2007). Empirical evidence on the determinants of team performance in Korean baseball league. *The Korean Journal of Measurement and Evaluation in Physical Education and Sport Science*, **9**, 63-77.
- MLB. http://mlb.mlb.com/stats/league_leaders.jsp.
- Park, S. H. (2008). Performance factors affecting the high annual salary of Korean professional baseball batters. *The Korea Journal of Sports Science*, **17**, 485-494.
- 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.

Batting index prediction model 2017

Chong Sun Hong¹ · Dong Sik Shin²

¹²Department of Statistics, Sungkyunkwan University

Received 3 April 2017, revised 8 May 2017, accepted 15 May 2017

Abstract

In this paper, we propose batting index prediction models of 2017. Due to the insufficiency of KBO pitchers data, batting index prediction models of 2016 has been developed based on elected eight batting index collecting the past three years data of MLB and KBO. It has been found that this prediction model fits well to both MLB and KBO, and the KBO model fits better than MLB in some cases. Using these prediction models, we analyzed and compared 2016's estimated values for the batting index of MLB and KBO. With the relation results between batting index prediction and batter's age for MLB and KBO, it can be determined that there is no relationship between the significant batting index and ages.

Keywords: Batting, index, longitudinal study, pitching, weighted mean.

¹ Corresponding author: Professor, Department of Statistics, Sungkyunkwan University, Seoul 03063, Korea. E-mail: cshong@skku.edu

² Graduate student, Department of Statistics, Sungkyunkwan University, Seoul 03063 Korea.