

프로야구 기록을 통한 승리 요인에 관한 연구*

김창권¹, 진서훈²

요 약

현대 스포츠는 하루가 다르게 발전하고, 그 규모 또한 날로 거대해 지고 있다. 특히 한국 프로야구는 남녀노소 누구나 좋아하는 스포츠로 자리 잡았다. 야구는 이미 스포츠를 넘어 하나의 문화로 자리 잡았고, 그 관중의 수도 매년 증가하는 추세이다. 야구는 통계와 밀접한 관련이 있는 스포츠로 통계의 활용도가 매우 높은 스포츠이며 각 팀들은 그 통계를 활용한다. 그렇지만 흔히 알고 있는 타율, 홈런, 삼진 등의 수치만이 팀의 승리를 결정하는 것은 아니다. 야구는 베이스를 얻어야 하는 운동이다. 야구가 베이스볼(baseball)이라고 불리는 이유이기도 하다. 베이스를 얼마나 얻었고, 잃었는가에 대한 문제가 가장 중요하다고 할 수 있다. 본 연구에서는 베이스 획득과 허용이 승리에 얼마나 영향을 주는지 알아보았다. 또한, 베이스 허용에 관해 우리가 잘 알고 있는 이닝 당 출루 허용수에 대해 문제점을 제시하고, 이닝 당 베이스 허용수란 통계수치를 만들었다. 회귀모형을 통해 출루율, 장타율과 더불어 이닝 당 베이스 허용수가 승패에 유의한 영향을 미치는 것을 확인하였다.

주요용어 : 이닝 당 출루 허용수, 장타율, 출루율, 회귀분석.

1. 서론

사회가 풍요로워짐에 따라 사람들은 여가를 즐기게 되었고, 그 중 스포츠 관람은 하나의 문화로 자리 잡게 되었다. 스포츠가 점점 우리의 생활과 밀접해지면서 스포츠 산업은 21세기 문화 산업의 중요한 인프라로 주목을 받고 있다(Lee, 1999). 특히 프로야구는 2006년 독일 월드컵의 여파로 그 관중의 수가 주춤하였지만 그 이후로 인기가 날이 갈수록 증가하고 있으며, 2010년에는 누적 관중 1억명을 돌파하였다. 또 매년 관중수가 늘어나고 있고, 2011년에는 600만명 이상이 경기장을 찾았다. 2012년에는 드디어 700만 관중을 돌파해 그 인기를 실감하고 있다(Figure 1). 야구(野球, baseball)는 누가 더 많은 루(base)를 얻었는가를 겨루는 경기이다. 홈에서 출발해서 1루-2루-3루를 거쳐 다시 홈으로 돌아오면 1점을 얻게 된다. 따라서 베이스볼(baseball)이라고 불린다. 베이스를 얻기 위해서 공격팀은 타격 등을 이용하여 출루를 하고, 수비팀은 그 출루를 막아내기 위해 노력한다. 즉, 승리는 어느 팀이 베이스를 더 얻는가, 그렇지 못하는가에 영향 받게 되어 있다.

야구는 시간제한이 없고 공평하게 27개의 아웃이 각 팀에게 주어지는 경기이다. 야수(野手)들은 공격과 수비, 그리고 주루 플레이를 잘하는 것이 필요 덕목으로 여겨진다. 하지만 우리나라의 프로야구는 공격에는 관대하나, 주루와 수비는 공격에 비해 관심이 덜 한 것이 사실이다. 물론, 홈런타

*본 연구는 고려대학교 특별연구비에 의하여 수행되었음.

*이 논문은 제1저자 김창권의 석사학위논문 일부 발췌한 것입니다.

¹339-700 세종특별자치시 연기군 서창리 208, 고려대학교 세종캠퍼스 과학기술대학 정보통계학과 대학원 석사과정. E-mail : fcerjtm@korea.ac.kr

²(교신저자) 339-700 세종특별자치시 연기군 서창리 208, 고려대학교 세종캠퍼스 과학기술대학 정보통계학과 교수. E-mail : seohoon@korea.ac.kr

[접수 2013년 11월 28일; 수정 2014년 1월 16일; 게재확정 2014년 1월 19일]

자들이 매력적인 것은 부정할 수 없으나, 야구는 팀 플레이어 각 위치에 맞는 선수들이 제 능력을 발휘할 때 많은 승리를 얻을 수 있는 경기이다. Shin, Park, Cho, Choi(2007)는 선수의 능력(성적)이 팀 승패에 주는 영향에 관하여 연구하였다. Kim(2013)은 팀의 기록과 팀의 승리에 관해 연구하였다. 본 연구에서는 홈런, 삼진 등의 기록이 아닌 베이스를 점유하는 기록에 관련하여 승리에 영향이 있는지 알아보도록 하겠다.

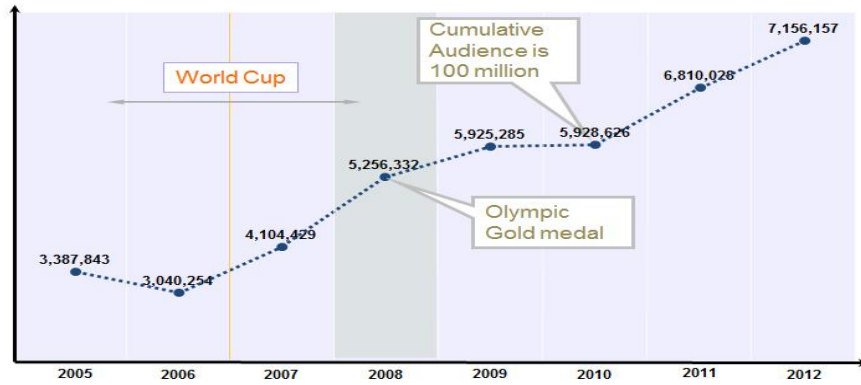


Figure 1. 2005~2012 Professional baseball attendance status

2. 야구에서의 통계 지수

2.1. 야구와 통계

야구는 통계의 스포츠라고 불린다. 타율, 방어율, 출루율 등의 통계 기록들이 널리 쓰이고 있다. 또 이런 기록들을 가공하여 사용하기도 한다. 기본적으로 야구에서 쓰이는 통계는 사칙연산을 통해 이루어진다. 야구에서의 통계는 투수가 공을 던지는 순간부터 심판이 볼 데드를 선언하기까지 모든 과정이 기록이 된다. 따라서 야구의 기록은 매우 복잡하며, 스코어북(야구 기록지)을 보면 그 경기가 눈앞에 펼쳐질 정도로 세밀하게 기록되어 있다.

2.2. 타자(batter, 打者) 관련 통계 지수

타자는 투수가 던진 공을 방망이로 쳐서 득점을 얻는 선수이다. 타자들은 타격과 타격 이외의 방법으로 베이스를 얻는다. 또한 더 많은 베이스를 얻기 위해 주루 플레이를 한다. 그리고 이들은 공격 이외에 수비도 한다. 이러한 타자들의 플레이에 대한 통계 지수를 살펴보면 다음과 같다.

1) 타율(batting average, AVG)

타자의 기록에는 여러 가지가 있지만, 그 중 타율이 가장 흔하게 접할 수 있는 기록일 것이다. 타율은 그 타자가 타석에 들어섰을 때 안타를 만들어내는 비율로서 식 (1)과 같이 총 안타 수(hits, H)를 타수(at-bats, AB)로 나누어준다.

$$AVG = H / AB \quad (1)$$

2) 출루율(on-base percentage, OBP)

출루율은 타자가 얼마나 베이스를 확보했는지에 대한 수치이다. 타자가 자신의 힘으로 베이스를

얻는 경우는 여러 가지가 있지만, 그 중 안타(H)와 볼넷(base on balls, BB)을 얻었을 경우 그리고 몸에 맞아서 베이스를 얻는 경우(hit by pitched, HP)를 출루율에 포함시킨다. 출루율의 공식은 다음의 식 (2)와 같다.

$$OBP = \frac{H + BB + HP}{AB + BB + HP + SF} \quad (2)$$

타율의 경우 안타로 인한 수치만을 나타내지만 출루율은 안타 이외의 여러 가지 경우를 고려하므로 타율보다 상대적으로 높다. 그렇지만 분모에 타수(AB), 볼넷과 몸에 맞는 볼 그리고 희생타(sacrifice flies and bunts, SF)의 경우를 고려하여 출루를 생각하기 때문에 그 차이가 크지는 않다. 통상적으로 출루율은 4할이 넘으면 매우 좋은 수치이다. 특히, 각 팀의 테이블 세터(팀의 1~2번 타자를 말함)는 이 출루율에 더욱 신경을 써야 한다. 이들 뒤에 포진하는 타자들은 주로 타격에 능력이 있는 선수들이기 때문에 득점에 더 높은 효과를 줄 수 있기 때문이다. 출루율은 타율과 다르게 타수로 나누어주는 것이 아니라, 타석($AB + BB + HP + SF$)으로 나누어준다. 출루는 안타가 아니더라도 여러 방법이 있기 때문에 모든 상황에서의 출루 비율을 나타내는 것이다.

3) 장타율(slugging percentage, SLG)과 공격공헌도(on-base percentage plus slugging percentage, OPS)

장타율은 타격을 통해 얼마나 많은 베이스를 얻었는가에 대한 수치이다. 즉, 단타는 1개의 베이스를 얻고 2루타는 2개, 3루타는 3개, 그리고 홈런은 총 4개의 베이스를 얻는 것이다. 장타율이 높다는 것은 그만큼 공을 멀리 보낼 수 있는 능력이 높다는 것이고, 장타율이 높은 선수들은 각 팀에 클린업 트리오(각 팀의 3번, 4번, 5번 타자를 말함)에 주로 배치된다. 장타율을 구하는 공식은 먼저 총 루타수(total bases, TB)를 구하여, 타수(AB)로 나누어주며 다음의 식 (3)으로 표현할 수 있다.

$$SLG = TB / AB \quad (3)$$

기본적으로 장타율은 5할 이상이면 아주 좋은 평가를 받는다. 출루율과 장타율, 이 두 가지를 합하여 만든 OPS라는 기록도 주요 기록 중 하나이다. OPS는 ‘공격공헌도’라고 불리며 $OBP + SLG$ 로 구할 수 있다. 이는 단순하지만, 타자의 타격 능력과 베이스 점유에 대한 좋은 지표를 제공한다(Lee, 2005).

2.3. 투수(pitcher, 投手) 관련 통계 지수

투수는 상대방 타자를 상대하기 위해 마운드에서 공을 던지는 선수이다. 그들은 상대가 치지 못하도록 포수와 사인을 주고받아 공을 던진다. 투수는 공을 던지는 순간부터 9번째 야수가 되며, 한국 프로야구에서는 타자가 될 의무는 없다. 투수와 관련된 통계 지수를 살펴보면 다음과 같다.

1) 평균자책점(earned run average, ERA)

평균자책점은 투수가 평균적으로 한 게임(9이닝 기준)당 허용한 점수를 의미한다. 실점과 관련된 통계이므로, 낮을수록 좋은 평가를 받는다. 방어율을 계산할 때에는 실점의 전부를 사용하는 것이 아니라 자책점(earned runs, ER)을 기준으로 한다. 자책점은 투수가 수비의 실책 등으로 출루시킨 주자가 득점을 한 경우는 포함되지 않고, 투수의 책임으로만 허용한 점수이다. 예를 들어 수비의 실책으로 주자가 한명 출루하고 다음 타자에게 홈런을 맞았다면, 실점은 2점이고 투수의 자책점은

1점이 되는 것이다. 이렇게 얻어진 자책점은 9를 곱하여 등판한 이닝의(innings pitched, IP) 수로 나누어준다.

$$ERA = \frac{ER \times 9}{IP}$$

이렇게 얻은 방어율은 소수점 아래 셋째 자리에서 반올림하여 사용한다. 투수의 경우 타자의 규정 타석과 성적이 같은 ‘규정 이닝’이 있다. 이는 소속팀이 소화한 경기의 수와 같다. 즉, 팀이 126 경기를 가졌다면 126이닝이 규정 이닝이 되는 것이다(Moon, 2009).

2) 이닝 당 출루 허용수(walks plus hits divided by innings pitched, WHIP)

이닝 당 출루 허용수는 투수가 이닝 당 몇 명의 주자를 허용했는가에 대한 수치이다. 안타 또는 볼넷을 얻어 출루를 한 주자가 몇 명인가 알아보는 수치인데, 1의 값을 가지면 상당한 수준으로 볼 수 있다. 그렇지만, 이 수치에는 문제점이 존재한다. 바로 장타(2루타 이상의 안타)에 대한 개념이 존재하지 않아 이 값만으로는 투수의 능력을 평가하기에 부족하다는 점이다. 다음의 표를 보자.

Table 1. Example of WHIP's problem

	innings	hits allowed	runs allowed	walks	WHIP
player A	9.0	9 home runs	9 runs	0	1.00
player B	9.0	7 hits	2 runs	3	1.11

선수 A의 경우 9이닝 동안 9개의 안타를 맞고, 볼넷을 하나도 주지 않아 이닝 당 출루 허용수는 $(9+0)/9=1$ 의 값을 얻게 된다. 선수 B의 경우는 7개의 안타와 3개의 볼넷을 허용했기 때문에 $(7+3)/9=1.11$ 의 이닝 당 출루 허용수를 얻었다. 그렇다면 과연 ‘선수 A가 더 좋은 투구를 보였다고 할 수 있는가?’라는 질문을 했을 때, 대답은 당연히 ‘아니다’이다. 당장 방어율만 보아도 선수 A는 9.00이고 선수 B는 2.00이다. 간단히 예를 들어 보아도 이닝 당 출루 허용수는 그 자체로 투수의 능력을 알아보기에 부족함이 있다. Lee, Bang(2010)과 Lee, Kim(2007)의 연구에서도 이닝 당 출루 허용수를 보완한 가중 이닝 당 출루 허용수(WWHIP)를 사용하여 승률을 예측하고 있다. 따라서 피장타의 개념과 사구와 같은 수치를 포함 시켜야 한다. 그래야 투수의 베이스 허용에 대한 설명을 더 잘 할 수 있다. 따라서 다음의 수치를 제안하도록 하겠다.

3) 이닝 당 베이스 허용수(allowed base by innings pitched, ABIP)

이닝 당 베이스 허용수는 말 그대로 한 이닝에 베이스를 얼마나 허용했는가에 대한 수치이다. 기존 이닝 당 출루 허용수를 보완한 것으로 장타(단타는 1개의 베이스, 2루타는 2개의 베이스, 3루타는 3개의 베이스 그리고 홈런은 4개의 베이스를 얻음)의 개념과 고의사구(intentional bases on balls, IBB)를 적용한 것이며, 구하는 식은 다음과 같이 표현 할 수 있다.

$$ABIP = \frac{(H + H2 \times 2 + H3 \times 3 + HR \times 4) + BB + HP + IBB}{IP} \quad (4)$$

Table 1에서 보면 선수 A의 경우 9개의 홈런을 허용했기 때문에 $(9 \times 4 + 0)/9 = 4$ 의 값을 가지고, 선수 B의 경우 모두 단타를 허용했기 때문에 이닝 당 출루 허용수와 같이 1.1의 이닝 당 베이스 허용수값을 가진다. 즉, 선수 A는 한 이닝동안 4개의 베이스를 허용하는 것이다. 베이스를 허용하는 관점에서 이닝 당 출루 허용수보다 이닝 당 베이스 허용수를 사용하는 것이 더 좋은 설명을 할 수

있을 것으로 보인다.

3. 실증 분석

3.1. 승리의 요소

프로 스포츠의 가장 큰 목적은 ‘승리’에 있다. 각 팀들은 승리를 얻기 위해 좋은 선수들을 모집하게 된다. 그렇다면, 각 프로야구 구단들이 승리를 위해서는 어떠한 요소에 중점적으로 투자, 훈련을 해야 할까?

기본적으로 야구는 득점을 많이 해야 하고, 실점을 적게 해야 많은 승리를 얻을 수 있다. 이 분야의 선행연구로 Cho, Cho(2005), Lee, Lee(2008)는 득점과 실점이 승률에 주는 영향에 관하여 분석하였다. 다음의 Figure 2는 2005년부터 2012년까지 모든 팀의 득점과 실점의 차이(가로축)가 승수(세로축)에 영향을 주고 있음을 잘 나타내 준다.

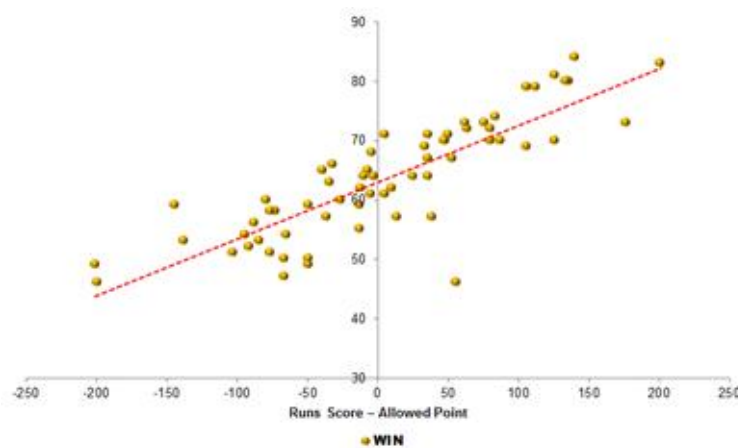


Figure 2. Relationship between win & runs allowed

1) 득점에 미치는 요소

득점을 하려면 4개의 베이스를 연속으로 얻어야 하고, 실점은 반대로 4개의 베이스를 연속으로 허용하면 된다. 따라서 득점을 위해 베이스를 우선적으로 얻어야 한다. 앞서 소개한 세 개의 변수(타율, 출루율, 장타율)와 득점이 얼마나 큰 상관이 있는지 먼저 알아보기 위해 상관분석(correlation analysis, 相關分析)을 실시하였다. 상관분석은 두 변량간의 상호관계를 측정하는 척도로서 사용된다.

Table 2. Correlation analysis of runs earned, batting average, on-base percentage, slugging average(N=64)

	R	AVG	SLG	OBP
R	1	-	-	-
AVG	0.79016 <.0001	1	-	-
SLG	0.85942 <.0001	0.76256 <.0001	1	-
OBP	0.81709 <.0001	0.81178 <.0001	0.76132 <.0001	1

이후에 이용된 데이터는 2005년부터 2012년까지의 각 팀별 기록으로 한국프로야구 홈페이지에서 수집하였다. 이 기간 동안 프로야구는 8개 구단으로 운영되었으므로 8개 년 8개 구단의 총 64개의 연 단위 데이터를 분석에 이용하였다.

타율(0.79), 출루율(0.82), 장타율(0.86) 모두 득점과 큰 양의 상관을 가지고 있는 것을 알 수 있다. 이는 타율, 장타율, 출루율이 높을수록 득점이 많아짐을 의미한다. 그리고 출루율과 장타율은 타율에 비해 더욱 높은 것을 알 수 있다. 이는 베이스의 선점이 득점에 많은 영향을 미칠 것이라고 예상하게 해준다.

2) 실점에 미치는 요소

득점과 마찬가지로 실점에 미치는 요소를 상관분석을 통해 알아보았다. 실점은 투수가 베이스를 허용하는 만큼 많이 이루어진다. 그래서 투수 기록 중 위에 제안한 이닝 당 베이스 허용수와 다음에 제시할 이닝 당 추가 진루 허용수(EBIP(extra bases by innings pitched))를 주요 요인으로 설정하였다. 이닝 당 추가 진루 허용수는 투수가 안타를 허용하는 것 이외에 야수실책, 와일드피치 등으로 추가 베이스를 주는 경우를 말한다. 앞서 이야기했듯이, 베이스를 얼마나 허용하느냐에 따라 그 팀의 실점을 가늠할 수 있다. 야수의 실책으로 추가적으로 베이스를 허용한다면 투수가 안타를 허용하는 경우와 마찬가지로 실점을 할 수 있다. 그렇기 때문에 수비적 요소를 가미한 추가베이스 허용에 대한 수치를 다음의 식 (5)와 같이 제안한다.

$$EBIP = \frac{E + WP + BK}{IP} \quad (5)$$

와일드 피치(wild pitches, WP)는 투수가 포수가 받기 어려운 곳으로 공을 던져 포수가 놓치는 경우를 말한다. 이 경우는 볼 데드 상황이 아닌 인 플레이(in-play)상황이기 때문에 주자가 있을 경우 자유롭게 플레이를 펼치게 된다. 또 부정투구(balks, BK)는 주자가 있을 때 주자를 기만한 경우로 주자에게 한 베이스를 안전하게 진루할 수 있게 하는 것이다. 또한 실책(E) 역시 아웃을 시킬 수 있는 상황에서 아웃을 시키지 못해 주자를 살리는 행위로 실점의 확률이 늘어나게 된다. 투수의 투구로 베이스를 허용하는 이닝 당 베이스 허용수의 수치에 투구 이외의 결과로 베이스를 추가로 허용하는 이닝 당 추가 진루 허용수를 합한 수치를 변수(총 베이스 허용률(OBR; on-base rate))로 사용하여 실점(P_R)에 대한 상관을 알아보면 다음과 같다.

Table 3. Correlation analysis of allowed points and OBR(N=64)

	P_R	OBR
P_R	1	-
OBR	0.90335 <.0001	1

Table 3에서 보듯 총 베이스 허용률(OBR)은 실점에 관하여 0.90의 상관을 보였다. 이는 베이스의 허용이 실점에 얼마나 중요한지 알 수 있는 결과이다. 이닝 당 출루 허용수의 경우 실점과 0.84의 상관을 보인다.

3.2. 베이스 점유, 허용과 승리와의 관계 모형

본 절에서는 실제 2005년부터 2012년까지의 팀 승리에 앞서서 살펴본 요인들이 미치는 영향에

대해서 알아보도록 하겠다. 앞서 소개한 베이스 선점에 관련된 5개의 변수(타율, 출루율, 장타율, 이닝 당 베이스 허용수, 이닝 당 추가 진루 허용수를 사용하여 승리에 각각의 수치들이 얼마나 영향력을 가지는지 알아보았다. 단, 회귀식에 사용된 관측치는 각 년도의 개체 간 독립성을 위하여 승률이 0.5에 가까운 한 팀을 제외한 7팀씩 8개년 총 56개를 사용하였다(7개 팀의 승률이 결정되면 나머지 한 팀의 승률은 자동 결정된다).

고려한 5개의 설명 변수 중에 타율(AVG)은 다른 변수들과의 상관성이 높은 것으로 판단하여 제외하였다. 또한 이닝 당 추가 진루 허용수는 승수에 유의한 영향을 미치지 못하였다. 따라서 두 변수를 제외한 다음의 모형을 만들었다.

$$Win = \beta_0 + \beta_1 SLG + \beta_2 OBP + \beta_3 ABIP + \epsilon \quad (6)$$

이 모형의 추정된 회귀 계수는 다음과 같다.

$$\widehat{WIN} = 6.45487 + 111.1594 \times SLG + 296.2463 \times OBP - 44.4122 \times ABIP \quad (7)$$

식(7)은 장타율이 1할 증가하면 +11.1승의 효과를, 출루율이 1할 증가하면 +29.6승, 그리고 이닝 당 베이스 허용수가 1할이 증가하면 -4.4승의 효과를 보이는 것으로 해석할 수 있다($R^2=0.72$).

Table 4. Regression parameter of final model

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	6.45487	21.27235	0.3	0.7628
SLG	1	111.1594	50.53403	2.2	0.0323
OBP	1	296.2467	90.80775	3.26	0.002
ABIP	1	-44.4122	4.97009	-8.94	<.0001

한편 이닝 당 출루 허용수를 이닝 당 베이스 허용수로 보완하지 않고 회귀식을 만든 Table 5의 결과를 살펴보면, 추정모수의 수치 차이는 있지만, 식 (7)에 크게 벗어나지 않는다. 그렇지만 $R^2=0.69$ 의 값을 보여 이닝 당 출루 허용수와 이닝 당 베이스 허용수의 실점에 대한 영향이 승리에 대한 설명력 차이를 보여준다.

Table 5. Regression parameter of final model with WHIP

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	47.70194	23.36945	2.04	0.0469
SLG	1	95.88326	49.38659	1.94	0.0576
OBP	1	253.61012	90.67145	2.80	0.072
WHIP	1	-75.48359	8.27503	-9.12	<.0001

모형에 사용된 각 변수들의 잔차분석 그림인 Figure 3을 보면 각 변수가 특별한 패턴을 보이지 않고 있음을 알 수 있다.

4. 결론

‘베이스를 얻는다.’는 가장 원초적인 문제로 각 팀의 승리 요소들을 알아보았다. 기존에 우리가 흔히 사용하던 홈런, 삼진과 같은 변수가 아닌 실제 베이스를 얻는 기준에서의 변수를 제시하고, 그 결과를 보였다. 야구의 기본은 모든 베이스를 돌아 홈으로 돌아와 득점을 하고, 상대방의 득점을 막는 것이다. 그래서 가장 단순하고도 중요한 베이스의 점유와 허용에 관해서 승리의 요소들

정했다. 그 요소는 팀 승리에 중요한 영향을 준다는 것을 분석을 통해 알 수 있었고, 공격과 수비 모두 좋은 역할을 할 때 상위권을 유지할 수 있다는 것을 알았다.

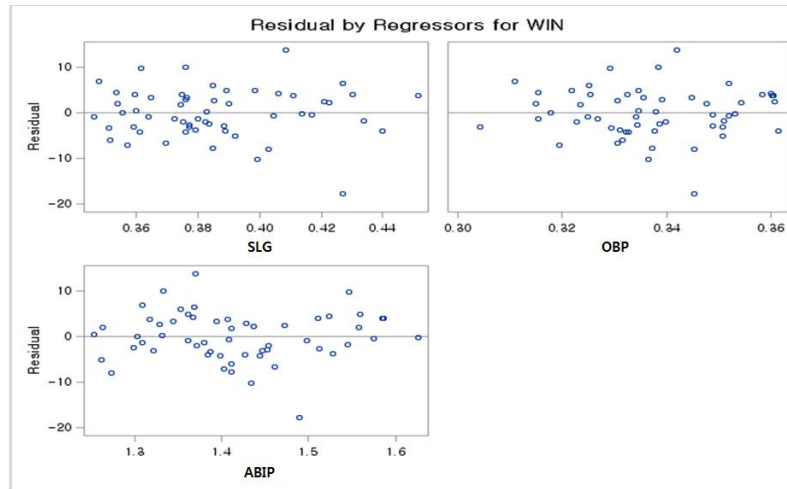


Figure 3. Residual analysis for each variable

흔히 타자의 평가는 타율에 의해 판단된다. 타율이 높다는 것은 많은 안타를 생산한다는 의미이다. 하지만 안타는 베이스를 얻기 위한 여러 방법 중 하나이다. 타율 이외에 출루율, 장타율 등이 승리에 영향을 미치는 요인이 됨을 분석을 통해 살펴보았다. 한편 투수의 평가에 관해서는 투수의 베이스 허용률이라는 값을 이닝 당 출루 허용수의 보안을 통해 얻을 수 있었다. 주루의 경우 도루와 도루 실패의 수치는 잘 관리되고 있다. 그러나 추가 진루와 견제사 그리고 주루사 등도 베이스를 얻는 중요한 요소이므로 관리가 필요하다. 수비의 경우도 마찬가지이다. 수비를 통해 베이스 허용을 막을 수 있음에 이 또한 중요한 요소인 것이다. 야구는 던지고 치는 것이 주된 목적이지만, 현대 야구는 이외에도 여러 가지의 작전을 구사하고 한 개의 베이스를 더 얻기 위해 각 팀들은 고민하고 행동한다. 따라서 앞으로는 이러한 자료들의 중요성을 인지하고, 프로야구 전반적으로 기록들을 인정해주는 분위기가 되어야 할 것이다.

References

- Cho, Y.-S., Cho, Y.-J. (2005). 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(6), 2303-2312. (in Korean).
- Kim, C. K. (2013). *A Study on the Winning Factors based on the Professional Baseball*, Master Thesis of the Department of Informational Statistics at Korea University. (in Korean).
- Lee, E. J. (2005). *Mathematician at the ballpark*, Whistler, 13-27. (translated in Korean).
- Lee, K. J. (1999). *Sports mathematical science*, Shinkwang-munwhasa, 1-2. (in Korean).
- Lee, J.-T., Bang, S.-Y. (2010). The Changing Team Performance in Korean Pro-Baseball, *Journal of the Korean Data Analysis Society*, 12(1B), 553-561. (in Korean).
- Lee, J.-T., Kim, Y.-T. (2007). An Effective Statistical Model that Predicted Winning Percentage in Korean Pro-Baseball, *Journal of the Korean Data Analysis Society*, 9(2), 931-942. (in Korean).
- Lee, H.-Y., Lee, S.-K. (2008). Relation Analysis Between Victory and the Records of Korean Professional Baseball, *Journal of the Korean Data Analysis Society*, 10(6B), 3413-3422. (in Korean).
- Moon, E. S. (2009). *Watching baseball smarter*, Bonus, 181-197. (translated in Korean).

- Shin, S.-K., Park, K.-S., Cho, Y.-S., Choi, S.-H. (2007). A Study on Analyzing Factors Affecting the Outcome of Korean Professional Baseball Games: A Case of Samsung Lions, *Journal of the Korean Data Analysis Society*, 9(4), 2071-2083. (in Korean).

Predicting Wins through Professional Baseball Statistics*

Chang Kwon Kim¹, Seohoon Jin²

Abstract

Professional sports are changing on a daily basis, and also the size of its market. Korean professional baseball established itself as a national pastime. Baseball is beyond a sport and now a national culture. The attendance and audience for the sport is increasing annually. Statistics plays a big role in baseball and its application is used by each team. However, conventional stats such as hitting percentage, home runs, strike-outs, and etc. will not predict a team's win. Baseball is a sport of earning runs through bases. This paper predicts win through understanding how bases are earned and given. This paper will also point out the flaws of WHIP, and present ABIP (allowed base by inning pitched) as a substitute. This paper shows on-base percentage, slugging percentage, and ABIP are significant factors to predict wins.

Keywords : WHIP, SLG, OBP, regression.

*This study was supported by Korea University Grant.

*This study is extracted from the 1st author's thesis for master degree.

¹Student in Master course, Dept. of Informational Statistics, Korea University, Sejong, 339-700, Korea.
E-mail : fcerjtm@korea.ac.kr

²(Corresponding Author) Professor, Dept. of Informational Statistics, Korea University, Sejong, 339-700, Korea. E-mail : seohoon@korea.ac.kr

[Received 28 November 2013; Revised 16 January 2014; Accepted 19 January 2014]