

시계열 분석을 통한 프로야구리그 경기력 예측

오승욱(경희대학교, 연구교수) · 한진욱*(경희대학교, 교수)

본 연구는 국내 프로야구 경기력 변화 트렌드를 파악하고 향후 경기력을 예측하는 데 그 목적이 있다. 이러한 연구목적을 달성하기 위해 'R' 프로그래밍 언어를 이용하여 ARIMA 시계열 분석을 실시하였으며, 이에 따른 결과 및 결론은 다음과 같다. 첫째, 홈런에 대한 경기력 변화예측은 지금까지 나타난 프로야구리그의 평균보다 약간 높은 수준을 보여주고 있다. 둘째, 타율에 대한 경기력 변화예측은 지금까지 나타난 국내 프로야구리그의 평균보다 매우 낮은 수준을 보여주고 있다. 셋째, 타점에 대한 경기력 변화예측은 지금까지 나타난 프로야구리그의 평균보다 높은 수준을 나타내고 있다. 본 연구 결과에 대한 인사이트 및 제언이 제시되었다.

※ 주요어 : 시계열 분석, ARIMA, 프로야구, 경기력, 예측

* 교신저자 : hjw5893@khu.ac.kr

서론

국내 프로야구는 2015년 10개 구단으로 리그를 확장하면서 지속적인 관중 수의 증가를 이어 오고 있다. 최근에는 정보통신 및 미디어 기술의 발달로 스포츠 분야의 데이터 활용이 많이 늘어나고 있으며, 특히 프로야구에서는 데이터를 경기 전략에 활용함으로써 경기운영의 효율성을 제고하고 있다(고동현, 박운성, 배원호, 홍석만, 2019). 미국 메이저리그 구단들은 데이터 전문가 영입을 통해 프로야구 데이터 분석을 진행하고 있으며, 선수 영입, 선수 훈련, 경기 전술, 선수 계약 등과 같이 다양한 영역에서 데이터 분석 기반의 구단 운영을 하고 있다(신호중, 2017).

야구에서는 경기상황 속에서 수많은 데이터가 쏟아져 나온다. 특히 프로야구 선수의 경기력에 대한 데이터는 많은 팬의 관심사가 되고 있다(한진욱, 안정찬, 오승욱, 신동일, 2015). 이에 방송 및 언론은 다양한 프로야구 경기력 지표를 노출하고 있고, 방송에서 노출되는 경기력 지표는 타율, 홈런, 안타, 도루 등의 전통적인 지표를 넘어서 MLB에서 제시되는 WAR, BABIP, FIP 등도 나타나게 되었다(Baumer & Zimbalist, 2014). 특히 대체 선수대비승수를 나타내는 WAR은 세이버메트릭스 통계량 중의 하나로 선수의 종합적인 가치를 표현하는 지표로서 시대별, 국가별 리그의 차이를 반영하여 나타냄으로써 선수 간 직접적인 비교가 가능하기에 그 효용성이 높다(Chang & Zenilman, 2013). 프로야구에서 경기력 지표는 팬과 언론에서뿐만 아니라 프로야구 현장에서도 구단의 선수관리 및 전략적 운영관리에 직접적으로 활용되고 있다. 프로야구단의 가치는 시장 가치, 구장의 가치, 브랜드 가치 등으로 구분하여 평가되는데, 소속구단 선수들은 이러한 가치의 향상에 많은 기여를 한다(송기성, 2000). 이에 구단들은 해당 구단의 가치를 높이기 위한 전략으로 선수의 가치로 평가되는 경기력 향상에 중요성을 인식하게 되었고, 그에 대한 전략적인 운영관리를 하고 있다(서덕원, 2008). 또한, 구단은 선수단 연봉의 과도한 증가에 따른 재정수지의 어려움을 극복하고자 객관적인 경기력 데이터를 기반으로 선수의 생산성 가치를 정확하게 측정하기 위한 노력을 하고 있으며, 이를 통해 상대적으로 연봉이 적은 선수들을 발굴하고 영입함으로

써 구단의 인적자원관리에 대한 효율성을 높이고 있다(Ng, 2017).

프로스포츠 리그의 가장 중요한 목표는 리그의 경쟁력 확보와 이윤 창출이다. 프로야구리그의 재정적 안정성을 확보하기 위해서는 프로리그 내 구단 간 경쟁균형과 리그 자체의 경기력 향상을 위한 노력이 요구된다. 프로스포츠 리그에서 관중 확보는 프로스포츠 리그의 존속과 성장에 절대적이며, 프로스포츠 리그는 관중 동원을 통한 직접적인 수익 창출뿐만 아니라 중계권 및 스폰서십 등의 막대한 파생 수익을 높일 수 있다(채한승, 이종호, 2000). 프로스포츠는 독점적 상품을 통해 경쟁력을 갖는 일반 산업의 특징과 다르게 리그의 경쟁력을 갖추어야 한다. 즉 리그에서 활동하는 프로구단 전체가 경쟁균형을 이루어야 프로스포츠 리그의 경쟁력을 높일 수 있다(Madrigal, 1995).

Rascher, McEvoy, Nagel과 Brown(2007)은 관중수요에 영향을 미치는 요인으로 경기력을 제시하였다. Howard와 Crompton(2004)은 상대 팀의 경기력을 경기관람의 중요 요인으로 인식하여 프로야구의 경기에서 상대 팀의 경기력이 높을수록 더 비싼 입장권을 받을 수 있음을 제시하였다. 장원용(2018)은 프로야구 관람 결정 요인 중 경기상황을 가장 중요한 요인으로 제시하면서 구체적으로 경기력의 향상을 강조하였다. 장진희와 문춘걸(2014)은 프로야구 구단의 전력 수준과 선수의 경기력 향상은 프로야구리그의 수준을 높인다고 주장하였다. 특히 이상학(2023)은 국내 프로야구의 심각한 거포 부재를 설명하면서, 야구의 꽃으로 평가받는 홈런을 칠 수 있는 타자의 부재가 프로야구리그 흥행에 악재가 될 수 있음을 심각하게 우려하고 있다.

이와 같이 프로스포츠 리그에서 경기력은 경기관람 및 관중수요에 중요한 요인이며, 경기력의 향상은 곧 프로스포츠 리그의 경쟁력을 확보할 수 있다. 이에 프로스포츠 리그를 담당하는 조직은 수시로 리그의 경기력을 확인하고 변화에 대응해야 할 필요성이 있다. 따라서 국내 프로야구리그를 관리하는 KBO는 리그의 경기력 변화를 파악하고 예측함으로써 리그의 성장자립도 및 경쟁력 강화를 위한 다양한 제도 및 정책이 요구되며, 이를 실현하기 위한 학문적 연구가 요구된다.

프로야구 구단 및 리그에서 경쟁력 확보를 위해 중요

하게 인식되는 경기력은 학계에서도 다양하게 연구되어 오고 있다. 국내 프로야구 시장규모의 확대 및 경기력 향상으로 인해 프로야구 선수의 가치는 매우 높은 수준으로 올라가고 있다. 국내 프로야구 선수들의 가치 증가는 과도한 연봉의 증가로 이어지고 있어 국내 프로야구 단들의 심각한 재정난을 초래하여 구단 운영에 어려움을 주고 있다. 이에 학계에서는 선수의 가치를 측정할 수 있는 경기력과 선수들의 받는 연봉 수준의 관계를 실증적으로 분석하는 연구가 주로 수행되고 있다(박승현, 2008; 송희배, 강기훈, 2012; 오태연, 이영훈, 2016; 이장영, 강효민, 2001). 특히 FA를 체결한 선수들은 고액의 연봉을 받는 만큼의 경기력을 보이지 않는다는 주장이 제기되면서(Stiroh, 2007), FA 선수를 대상으로 경기력과 연봉 수준의 관계를 실증적으로 분석하는 연구가 수행되고 있다(홍종선, 원치환, 정동길, 2015). 지금까지 수행된 연구는 선수의 가치와 연봉의 합리성을 설명함에 있어 선수의 경기력 요인을 활용하였으며, 이를 통해 프로야구 구단의 효율적인 운영을 위한 전략을 제시하였다. 하지만 경기력은 선수의 가치를 평가하는 요인임과 동시에 프로야구리그의 가치를 평가하는 중요 요인임에도 불구하고 그에 대한 연구는 전무한 실정이다. 경기력은 프로스포츠 리그의 경쟁력을 위한 중요 요인으로 인식되는 만큼, 과거의 프로스포츠리그 경기력 변화를 파악하고 미래 경기력을 예측하는 연구가 필요해 보인다.

일반적으로 기존의 데이터를 바탕으로 예측하는 방법으로는 시계열 데이터 분석이 주로 활용되고 있다. 최재일과 정용락(2010)은 한국프로축구 관중 수를 예측하는 연구에서 시계열 분석을 활용하였고, 설민신, 박두용과 이미정(2011)은 2010년까지의 프로야구 관중 수 데이터를 기반으로 시계열 데이터 분석을 활용하여 2011년에서 2015년까지 약 5년 동안의 향후 관중 수를 예측하는 연구를 수행하였다. 시계열 분석을 통한 데이터 예측은 미래의 상황을 미리 파악함으로써 어떤 문제를 효과적으로 대응할 수 있게 한다.

따라서 본 연구는 국내 프로야구 경기력 변화 트렌드를 파악하고 향후 경기력을 예측하는 데 그 목적이 있다. 이를 위해 국내 프로야구가 창설된 1982년부터 2022년까지 약 41년간의 경기력 데이터를 수집하여 시

계열 ARIMA 분석을 통해 규명하고자 한다. 선행연구에서 살펴본 바와 같이 경기력은 프로야구 경기의 중요 경기관람 요인으로 프로야구리그의 관중 동원에 큰 효과를 나타낼 것이며, 나아가 중계권 및 스폰서십 등과 같은 파생 수익 창출에도 기여할 수 있다. 프로야구리그 경기력은 한 시즌이 끝나고 경기력이 어땠는지를 평가할 수 있는 사항으로 만약 선제적으로 변화를 예측하여 대응한다면 리그의 경쟁력 확보를 위한 중요 수단으로 활용될 수 있다. 즉, 프로야구리그의 경기력 변화 예측이 과거의 데이터에 비해 경기력이 감소할 것으로 나타난다면, KBO는 프로야구리그의 경기력 향상을 위한 전략을 미리 수립할 수 있을 것이다. 본 연구 결과는 프로야구리그의 경쟁력 확보를 위한 정책 및 제도를 선제적으로 수립하는데 중요한 기초자료가 될 것으로 판단된다. 또한 프로야구 구단들이 선수 영입에 있어 효율적 운영을 위한 기초자료로 활용할 것으로 기대된다.

연구방법

1. 자료 수집

본 연구에서는 국내 프로야구리그 경기력을 파악하고 향후 경기력을 예측하고자 년 도별 프로야구 경기력 데이터를 수집하였다. 데이터 수집기간은 프로야구 출범 이후부터 2022년까지 41년간으로 선정하였으며, KBO 홈페이지 기록 카테고리에서 해당 년도에 규정타석을 채운 선수에 대해서 경기력 데이터를 수집하였다. KBO 홈페이지에서는 타율, 경기 수, 타석 수, 타수, 안타, 2루타, 3루타, 홈런, 타점, 도루 등 열 개의 데이터를 타자의 경기력 지표로써 1982년부터 2022년까지 공통으로 제시하고 있다. 이에 본 연구에서는 경기력을 파악하기 위한 경기력 지표로 타격의 3대 지표이자 전통적으로 가장 널리 쓰이고 있는 타율, 홈런, 타점 등 총 세 개의 지표를 선정하였다.

2. 연구절차 및 방법

본 연구에서는 국내 프로야구리그의 년도별 경기력 지표를 바탕으로 향후 국내 프로야구리그의 경기력을 예측하는 데 목적이 있다. 연구를 수행하기 위한 구체적인 연구절차는 <그림 1>과 같다.

첫째, 국내 프로야구리그의 경기력 데이터를 수집하기 위해 KBO(<https://www.koreabaseball.com/>) 홈페이지를 활용하였다.

둘째, 수집된 자료는 본 연구의 목적에 적합하도록 정제작업을 수행하였다. 본 연구는 년도 별 경기력 데이터가 요구되기에 선수별 경기력 데이터를 해당 년도 데이터로 통합하는 과정을 거쳤으며, 년도 별 경기 수의 차이를 반영하여 2022년 144경기를 기준으로 경기력 지표 값을 표준화하였다. 또한 본 연구에 활용할 경기력 지표(타율, 홈런, 타점)만을 추출하였다.

셋째, 본 연구의 목적인 국내 프로야구리그의 경기력을 예측하기 위해 ARIMA 시계열 분석을 활용하였으며, 'R' 프로그래밍 언어를 사용하여 분석하였다.

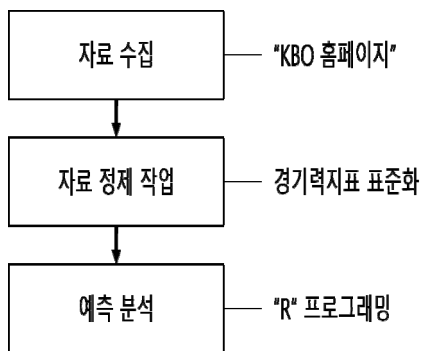


그림 1. 연구절차

1) 시계열 분석

본 연구에 수집된 자료는 'R' 프로그래밍 언어를 이용하여 ARIMA 시계열 분석을 실시하였다. 시계열 분석은 다음과 같은 절차에 따라 진행된다.

첫째, 시계열 분석을 실행하기 위해서는 본 시계열 자료가 정상성(stationary)을 갖고 있어야 한다. 이에 대한 시계열 데이터의 정상성 확인은 Augmented Dickey

Fuller 패키지를 활용하여 진행하였다. ADF는 시계열 ARIMA 모형의 적분차수를 판별하기 위한 검정방법으로 주로 활용된다. 시계열 데이터가 정상성을 확보하지 못한다면 log 변환 및 차분(differencing)을 통해서 정상성을 확보할 수 있다. 본 연구에서는 차분을 통해서 시계열 데이터의 정상성을 확보하였다.

둘째, 시계열 데이터에 대한 시계열 ARIMA 모형을 적용하고자 한다면 시계열 데이터 값 간의 상관계수를 산출하여 ARIMA(p, d, q)모형을 구성한다. p값은 자기회귀를 나타내는 것으로 과거 관측값의 선형결합이 미래의 관측값에 영향을 준다는 의미이다. d는 시계열 자료에 대한 차분을 나타낸다. q는 이동평균을 나타내는 값으로 관측값은 과거의 오차항의 영향을 받는다는 의미이다. 일반적으로 ARIMA(p, d, q)모형은 자기상관함수(ACF)와 편자기상관함수(PACF)의 그래프를 분석하여 적절한 값의 p, q값을 정한다. 하지만 'R' 프로그래밍에서는 AUTO ARIMA 패키지를 활용하여 가장 적절한 p, q값을 구할 수 있다. 'R' 프로그래밍에서 auto.arima() 함수는 Hyndman-Khandakar 알고리즘을 통해 가장 작은 AICc로 최적의 모형을 선택한다. 만약 d=0이라면, 상수 c는 포함이 된다; 만약 d≥1이라면 상수 c는 0으로 설정된다. 현재 모델의 분산은 현재 모델에서 ±1로 p 또는 q를 변화시키고 현재 모델에서 c를 포함/제외함으로써 고려된다. 이 과정은 더 낮은 AIC를 찾을 수 없을 때까지 반복 된다.

셋째, 적절한 ARIMA(p, d, q)모형을 활용하여 향후 경기력의 변화를 예측하였으며, 'R' 프로그래밍의 "forecast" 패키지를 활용하였다. 즉 시계열예측은 과거의 데이터를 기반으로 트렌드를 파악하고 트렌드 변화의 특성을 확인하여 미래의 그 현상이 지속될 것이라는 가정하에 향후 데이터의 변화를 예측하는 것이다.

연구결과 및 논의

1. 시계열 정상성 확인

국내 프로야구리그의 경기력 지표의 시계열 데이터에

대한 정상성을 확인하고자 Augmented Dickey Fuller 패키지를 활용하였다. 국내 프로야구리그의 경기력 지표인 홈런, 타율, 타점의 세 개의 지표에 대한 시계열 정상성을 검증하였으며, 그에 대한 구체적인 결과는 <그림 2>와 같다. adf_check 함수를 활용하여 원 시계열 데이터에 대한 정상성을 검증한 결과 홈런 및 타율에 대해서는 정상성이 확보되었다($p=0.020$, $p=0.012$). 타점에 대해서는 정상성을 확보하기 위해 1차 차분을 실시하였으며, 이를 통해 정상성을 확보하였다($p=1.3546 \times 10^{-17}$).

```
from statsmodels.tsa.stattools import adfuller
def adf_check(ts):
    result = adfuller(ts)
    if result[1]<0.05:
        print('stationary{}'.format(result[1]))
    else:
        print('non-stationary{}'.format(result[1]))

adf_check(df1["hitrates"].dropna())
adf_check(df1["homeruns"].dropna())
adf_check(df1["runs"].dropna())

stationary0.02072428285826712
stationary0.012251164632094852
non-stationary0.5111622786333537

df1["runs 1st diff"] = df1["runs"] - df1["runs"].shift(1)
adf_check(df1["runs 1st diff"].dropna())

stationary1.3546957787054896e-17
```

그림 2. 시계열 데이터 정상성 확인

2. 홈런 경기력 변화 예측

본 연구에서는 프로야구 경기력 지표인 홈런에 대한 경기력의 변화를 예측하고자 시계열 모형을 활용하였다. 첫째, 홈런에 대한 경기력 시계열 데이터 자료의 정상성을 확인하였다. 둘째, 홈런에 대한 경기력 변화를 예측하기 위해 최적의 ARIMA 모델을 실현하고자 “R” 프로그래밍에서 제공하는 auto.arima() 함수를 적용하였으며, 그에 대한 결과는 <그림 3>과 같다. 본 연구에서는 최적의 ARIMA 모델로 ARIMA(0, 1, 1)를 적용하였으며, <그림 3>에서 제시된 $p<0.05$ 로 나타나 모델의 적합성을

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.6158	0.103	-5.969	0.000	-0.818	-0.414
sigma2	1.818e+04	3888.045	4.677	0.000	1.06e+04	2.58e+04
Ljung-Box (L1) (Q):			0.10		Jarque-Bera (JB):	0.64
Prob(Q):			0.75		Prob(JB):	0.73
Heteroskedasticity (H):			3.32		Skew:	0.28
Prob(H) (two-sided):			0.04		Kurtosis:	3.28

그림 3. 홈런 ARIMA 모델 적합성(0, 1, 1)

확인하였다. 홈런의 원 시계열 데이터의 정상성을 확인하였지만, 최적의 모델을 적용하기 위해 1차 차분한 시계열 데이터를 사용하였다. 셋째, 홈런에 대한 경기력 변화를 예측하기 위해 “R” 프로그래밍의 forecast R 패키지에 있는 forecast.Arima() 함수를 활용하였으며, 그에 대한 구체적인 값 및 추세 그림은 <표 1>, <그림 4>와 같다. 구체적으로 살펴보면, 예측값은 2023년부터 2027년까지 모두 같은 620.5개의 홈런을 기록할 것으로 나타났다. 이는 과거 41년 동안 프로야구리그에서 기록한 홈런 수의 평균인 528.2개보다는 높은 수준이며, 전체 41년 간의 리그에서 상위 약 11위 수준의 경기력을 나타내고 있다(<표 2>). KBO 관중 현황 자료에 따르면, 2015년(7,360,530명)에 처음으로 700만 명을 돌파하였으며 이후 2018년까지 관중의 수가 역대 최고치를 경신하였다(2016년: 8,339,577명, 2017년: 8,400,688명, 2018년: 8,073,742명). <그림 4>에 나타난 홈런에 대한 트렌드를 살펴보면 2014년 이후로 2018년까지 역대 가장 많은 홈런을 기록하고 있는 것으로 보인다(2014년: 812.2개, 2015년: 705개, 2016년: 778개, 2017년: 689개, 2018년: 1,036개). 이는 홈런 수가 프로야구 관중수에 간접적인 영향을 미친 것으로 유추해볼 수 있다. 장원용(2018)은 남성에 있어 프로야구 관람 결정요인 중 경기력을 가장 중요한 요인으로 제시하고 있어 관중과 경기력 간의 상관관계를 판단할 수 있으며, 이상학(2023)은 홈런의 부재가 프로야구리그의 흥행에 영향을 미치고 있음을 제시하고 있다. 또한, 배정섭(2023)은 머신러닝을 활용한 플레이오프 진출구단의 타격 요인을 분석하는 연구에서 경기력 지표 중 홈런이 구단의 플레이오프 진출 여부에 유의미한 차이를 보인다고 제시하였다. 구단의 성적에 있어 홈런은 중요한 요인임을 알 수 있

표 1. 홈런에 대한 경기력 예측 value

	point .Forecast	Low	High
2023	620.5	447.7	793.3
2024	620.5	435.2	805.7
2025	620.5	423.6	817.4
2026	620.5	412.6	828.3
2027	620.5	402.1	838.8

Forecasts from ARIMA(0,1,1)

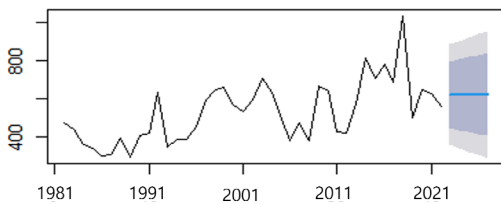


그림 4. 홈런 경기력 예측 plot

다. 이는 홈런에 대한 경기력이 구단의 가치뿐만 아니라 프로스포츠 리그의 경쟁력에도 영향을 주고 있음을 재확인할 수 있다.

2019년에는 홈런 수가 급감하여 499개를 기록하였는데 지속적인 관중 수 증가 추세에 부정적인 영향을 미친 것으로 판단된다. 이러한 영향은 공인구의 반발계수 합격 기준을 2019년에 약 0.1 낮아진 0.4034~0.4234로 변경하였기 때문인 것으로 파악된다. 공인구의 반발계수는 0.001에 따라 비거리에서 20cm 정도의 차이를 보이는 것으로 나타나 2018년 이후로 국내 프로야구리그의 홈런 수가 감소하고 있는 것으로 파악된다(김양희, 2022). 이러한 결과는 2023년 이후 예측된 홈런 수의 개수가 공인구 반발계수 변경 전인 2018년에 비해 낮게 나타나고 있어 향후에도 영향을 미칠 것으로 판단된다. 이에 국내 프로야구리그의 홈런 경기력을 향상시킬 수 있는 방안으로 공인구 반발계수의 변화를 고려해 볼 필요성이 있다.

3. 타율 경기력 변화 예측

본 연구에서는 프로야구 경기력 지표인 타율에 대한 경기력의 변화를 예측하고자 시계열 모형을 활용하였다.

표 2. 홈런 기술통계

index	홈런
count	41.0
mean	528.2
std	161.2
min	291.6
25%	393.3
50%	509.7
75%	636.6
max	1,036.0

첫째, 타율에 대한 경기력 시계열 데이터 자료는 정상성을 확인하였다. 둘째, 타율에 대한 경기력 변화를 예측하기 위해 최적의 ARIMA 모델을 실현하고자 “R” 프로그래밍에서 제공하는 `auto.arima()` 함수를 적용하였으며, 그에 대한 결과는 <그림 5>와 같다.

본 연구에서는 최적의 ARIMA 모델로 ARIMA(1, 0, 0)를 적용하였으며, <그림 5>에서 제시된 p-value <.05로 나타나 모델의 적합성을 확인하였다. 셋째, 타율에 대한 경기력 변화를 예측하기 위해 “R” 프로그래밍의 forecast R 패키지에 있는 `forecast.Arima()` 함수를 활용하였으며, 그에 대한 구체적인 값 및 추세 그림은 <표 3>, <그림 6>과 같다. 구체적으로 살펴보면, 예측값은 2023부터 2027년까지 0.282, 0.284, 0.285, 0.285, 0.286을 기록할 것으로 나타났다. 이는 과거 41년 동안 프로야구리그에서 기록한 타율의 평균인 0.287보다 낮은 값으로 나타나 향후 이에 대한 대비가 필요해 보인다(<표 4>). Hansen과 Gauthier(1989)는 프로스포츠 관전에서 관객 동원에 가장 영향을 미치는 요인으로 경쟁팀과 비교하여 드러나는 팀의 수준이라고 제시하였다. 설수영(2022)은 공격력 지표인 OPS가 한국 프로야구리그 구단의 경쟁력에 유의미한 영향을 미친다고 제시하였다. 공

	coef	std err	z	P> z	[0.025	0.975]
intercept	0.1218	0.039	3.148	0.002	0.046	0.198
ar.L1	0.5746	0.134	4.292	0.000	0.312	0.837
sigma2	6.67e-05	1.4e-05	4.759	0.000	3.92e-05	9.42e-05
Ljung-Box (L1) (Q):	0.34	Jarque-Bera (JB):	2.42			
Prob(Q):	0.56	Prob(JB):	0.30			
Heteroskedasticity (H):	1.92	Skew:	0.54			
Prob(H) (two-sided):	0.23	Kurtosis:	3.51			

그림 5. 타율 ARIMA 모델 적합성(1,0,0)

표 3. 타율 경기력 변화 예측

년도	point Forecast	Low	High
2023	0.282	0.271	0.292
2024	0.284	0.272	0.296
2025	0.285	0.272	0.297
2026	0.285	0.273	0.298
2027	0.286	0.273	0.298

Forecasts from ARIMA(1,0,0) with non-zero mean

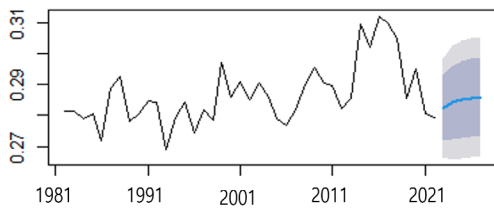


그림 6. 타율 경기력 변화 PLOT

경기력 지표인 OPS는 출루율과 장타율의 합으로 계산되며, 타율은 출루에 가장 많은 영향을 주는 요인으로 나타났다(조선미, 김주학, 강지연, 김상균, 2023). 이처럼 타율은 프로야구리그의 경쟁력 향상을 위해 필요한 요인으로 나타났다. 따라서 향후 타율의 경기력 감소로 나타난 본 연구의 결과는 향후 관중의 감소 및 리그의 경쟁력 약화로 이어질 상황을 예견한다. 조형석(2021)은 볼과 스트라이크를 구별하는 선구안보다는 정확하고 강하게 때려내는 것이 안타를 높일 수 있다고 제시하였다. 이에 국내 프로야구 선수들의 적극적인 스윙과 강한 타구 생산을 통해 향후 국내 프로야구리그의 안타 수 감소를 극복할 수 있을 것으로 판단된다.

본 연구에서 수집한 년도 별 선수의 경기력 데이터에 따르면, 프로야구 선수 중 41년 동안 규정타석을 채운 선수는 총 352명으로 나타났다. 그 중 9시즌 이상 규정타석을 채운 선수는 35명으로 나타나 약 10% 밖에 되지 않는다. 이는 경기력이 우수한 선수가 선수의 연령의 증가와 부상으로 인해 규정타석을 채우지 못함으로써 프로야구리그의 전체 경기력 하락으로 나타난 것이다. 오덕자와 박병주(2003)는 중고등 야구 선수의 타율과 체력 요인 관계를 살펴본 연구에서 타율은 근력, 순발력, 유연성의 요인과 유의한 관계가 있음을 제시하였다. 이처럼

표 4. 타율 기술통계

index	타율
count	41
mean	0.287
std	0.010
min	0.269
25%	0.281
50%	0.285
75%	0.291
max	0.312

경기력은 신체적, 기술적, 정신적 등의 총체적인 상호작용의 결과로 나타나는 것으로 향후 프로스포츠 리그의 경기력 향상을 위해서는 부상 방지를 위한 시설 및 정책의 필요성이 요구된다.

4. 타점 경기력 변화 예측

본 연구에서는 프로야구 경기력 지표인 타점에 대한 경기력의 변화를 예측하고자 시계열 모형을 활용하였다. 첫째, 타점에 대한 경기력 시계열 데이터 자료는 1차 차분을 통해 정상성을 확인하였다. 둘째, 타점에 대한 경기력 변화를 예측하기 위해 최적의 ARIMA 모델을 실현하고자 “R” 프로그래밍에서 제공하는 auto.arima() 함수를 적용하였으며, 그에 대한 결과는 <그림 7>과 같다. 본 연구에서는 최적의 ARIMA 모델로 ARIMA(0, 1, 1)를 적용하였으며, <그림 7>에서 제시된 p-value < .05로 나타나 모델의 적합성을 확인하였다. 셋째, 타점에 대한 경기력 변화를 예측하기 위해 “R” 프로그래밍의 forecast R 패키지에 있는 forecast.Arima() 함수를 활용하였으며, 그에 대한 구체

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.5915	0.139	-4.268	0.000	-0.863	-0.320
sigma2	1.756e+05	3.65e+04	4.807	0.000	1.04e+05	2.47e+05
Ljung-Box (L1) (Q):	0.02	Jarque-Bera (JB):	3.82			
Prob(Q):	0.88	Prob(JB):	0.15			
Heteroskedasticity (H):	3.71	Skew:	0.69			
Prob(H) (two-sided):	0.02	Kurtosis:	3.61			

그림 7. 타점 ARIMA 모델 적합성 (0.1.1)

표 5. 타점 경기력 변화 예측

년도	Point.Forecast	Low	High
2023	3,276.8	2,737.5	3,816.0
2024	3,276.8	2,691.8	3,861.7
2025	3,276.8	2,649.4	3,904.1
2026	3,276.8	2,609.7	3,943.8
2027	3,276.8	2,572.2	3,981.3

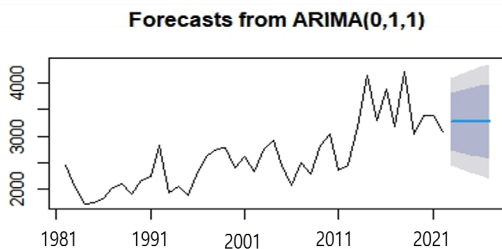


그림 8. 타점 경기력 변화 PLOT

적인 값 및 추세 그림은 <표 5>, <그림 8>과 같다. 구체적으로 살펴보면, 예측값은 2023년부터 2027년까지 모두 같은 3,276.5개의 타점을 기록할 것으로 나타났다. 이는 과거 41년 동안 프로야구리그에서 기록한 타점 수의 평균인 2,6182개 보다 높은 수준으로, 프로야구리그 전체 중에서 높은 수준을 나타내고 있다(표 6).

프로야구의 특성상 많은 득점을 해야 상대와의 경기에서 승리할 수 있기에, 타점은 매우 중요한 승리요인인 동시에 팬의 흥미를 유발하는 요인이기도 하다. 축구에 펠레스코어가 있다면 야구에는 케네디 스코어가 있다. 케네디 스코어는 프로야구 경기 스코어가 8:7인 경기가 가장 재밌다는 표현으로 KBO가 발행하는 기록집에도 나타나고 있다. 즉 프로야구 경기에서의 많은 득점을 유발하는 타점은 팬의 경기관람에 흥미를 유발하는 요인이다. 따라서 향후 타점에 대한 경기력이 높은 수준을 나타낼 것으로 예측된 본 연구결과는 미래에도 국내 프로야구리그의 경쟁력을 유지할 수 있을 것으로 판단된다.

결론 및 제언

본 연구는 국내 프로야구 경기력 변화 트렌드를 파악

표 6. 타점 기술 통계

index	타점
count	41
mean	2,618.0
std	623.9
min	1,733.8
25%	2,101.3
50%	2,458.3
75%	3,031.0
max	4,218.0

하여 경기력 변화를 예측하고, 향후 프로스포츠 리그의 경쟁력 향상을 위한 시사점을 도출하고자 하였다. 프로스포츠 리그에서 경기력은 리그의 생존과 자립에 있어 매우 중요한 지표로 나타났다. 이에 본 연구는 41년간 국내 프로야구리그의 경기력 데이터에 대한 변화추이를 살펴보고, 향후 경기력 변화를 예측하였으며, 그에 대한 시사점 및 제언은 다음과 같다.

첫째, 홈런에 대한 경기력 변화예측은 지금까지 나타난 프로야구리그의 평균보다 약간 높은 수준을 보여주고 있다. 경기력 지표 중 홈런은 프로야구리그의 관중 확보를 위한 중요 수단으로 여겨지는 만큼, 홈런에 영향을 주는 요인을 파악하여 추출된 요인에 대한 세부적으로 향상시킬 수 있는 전략이 하나의 방안이 될 수 있다.

둘째, 타율에 대한 경기력 변화예측은 지금까지 나타난 국내 프로야구리그의 평균보다 매우 낮은 수준을 보여주고 있다. 타율은 득점을 위한 중요 요인으로 구단의 경쟁력뿐만 아니라 스포츠 리그의 경쟁력에도 영향을 미친다는 것을 확인하였다. 이에 미래의 타율에 대한 경기력 감소에 대비하기 위해서는 국내 프로야구리그에서 활약하는 뛰어난 선수들의 경기참여를 높이는 방안이 필요해 보인다.

셋째, 타점에 대한 경기력 변화예측은 지금까지 나타난 프로야구리그의 평균보다 높은 수준을 나타내고 있다. 본 연구의 타점에 대한 결과가 높은 수준의 경기력 변화를 예측한 만큼, 향후 국내 프로야구리그의 경쟁력에 긍정적인 신호로 판단된다.

종합적으로 살펴보면, 홈런 및 타율에 대해서는 예측된 경기력보다 향상시킬 필요성이 요구된다. 배정성, 지준희와 원도연(2018)은 고교야구에서 홈런을 칠 있는 선

수가 부족한 것이 나무 배트의 사용으로 인한 것으로 제시하였다. 결과적으로 나무배트의 사용은 스윙의 변화로 인한 타격지표의 하락을 가져왔으며, 이는 프로진출 후에도 홈런의 감소로 이어지고 있다. 권훈(1999)은 외국인 선수 영입으로 인한 국내 프로야구 타자의 능력이 향상되었다고 밝혔다. 이는 프로야구 리그의 타자의 경기력 향상을 위한 수단으로 외국인 선수 중 타자의 보유 한도를 확대할 필요성이 존재한다(김근한, 2019). Schulz, Musa, Staszewski와 Siegler(1994)은 선수의 경기력이 가장 정점을 기록하는 연령에 대한 연구에서 타자는 주로 27~30세 사이에 가장 경기력이 정점에 있음을 밝히고 있다. 우리나라는 가장 정점에 있는 나이에 군 문제를 해결해야 되는 특수 환경에 놓여 있다. 이처럼 프로스포츠 리그의 경기력을 향상시키기 위해서는 경기력과 관련한 전체 환경요인을 분석하고 그에 대한 대처방안이 요구되는 만큼, 국내 프로야구리그의 경기력 감소의 원인으로 추측되는 고등야구의 알루미늄 배트의 재사용은 고려될만하며, 경기력 향상에 도움이 되는 외국인선수 제도는 타자 보유한도에 대해서 정책적으로 고려될 만한 것으로 판단된다. 또한 프로야구 선수의 군복무에 대한 역할 및 시기, 제도에 대해서도 많은 논의가 요구된다.

본 연구에서 진행한 프로야구 리그의 경기력 변화 예측은 41년간의 데이터를 바탕으로 제시했다는 점에서 그 의의가 있지만 다음과 같은 한계도 존재하여 그에 대한 제언을 하고자 한다.

본 연구에서는 전통적으로 활용되는 홈런, 타율, 타점에 대한 경기력 변화 트렌드를 살펴보았지만 프로야구리그에서 활용되는 다양한 경기력 지표가 있는 만큼, 세이버 매트릭스 경기력 지표를 추가한 연구도 필요해 보인다. 이를 수행한다면, 국내 프로야구리그의 경기력 변화를 종합적으로 판단할 수 있을 것으로 판단된다.

참고문헌

고동현, 박운성, 배원호, 홍석만 (2019). **수학을 품은 야구공**. 서울: 영진닷컴.

- 권훈 (1999. 10. 31). 프로야구 중흥 원년 절반의 성공. **연합뉴스** <https://sports.news.naver.com/general/news/read.nhn?oid=001&aid=0004478064>.
- 김근한 (2019. 1. 2). 속전속결 이끈 외국인 상한제, 보유 한도도 늘릴까. **엠스플뉴스** http://www.mbcsporsplus.com/news/?mode=view&cate=1&b_idx=99868002.000.
- 김양희 (2022. 3. 31). 프로야구 공인구 반발계수 5년 내 최소화...홈런주의보 없다? **한겨레**. <https://www.hani.co.kr/arti/sports/baseball/1037084.html>.
- 박승현 (2008). 한국프로야구 타자의 고액연봉에 영향을 미치는 경기력 요인. **한국체육과학회지**, 17(2), 485-494.
- 배정섭 (2023). 머신러닝을 활용한 플레이오프 진출을 위한 한국프로야구구단의 타격변수 예측. **한국체육과학회지**, 62(4), 155-163.
- 배정성, 지준희, 원도연 (2018). 고교야구에서의 나무배트 사용의 문제점에 따른 대안 도출. **한국레저사이언스학회지**, 9(1), 1-12.
- 서덕원 (2008). **프로스포츠선수의 근로자성에 관한 연구: 프로야구 및 프로축구 선수를 중심으로**. 미간행 석사학위 논문, 고려대학교 노동대학원.
- 설민신, 박두용, 이미정 (2011). 시계열 분석을 이용한 한국 프로야구 관중 예측연구 (2011-2015). **한국사회체육학회지**, 45(1), 375-387.
- 설수영 (2022). 한국 프로야구리그 구단의 경쟁력 요인 분석. **한국스포츠학회지**, 20(4), 583-592.
- 송기성 (2000). 독점규제 및 공정거래에 관한 법률이 프로스포츠 제도에 미치는 영향. **한국스포츠산업경영학회지**, 5(1), 17-27.
- 승희배, 강기훈 (2012). 한국 프로야구 선수들의 경기력과 연봉의 관계 분석. **한국데이터정보과학회지**, 23(2), 285-298.
- 신호중 (2017). **테오 애플스타인에게 배우는 33역량**. 파주: 넥서스BIZ.
- 오덕자, 박병주 (2003). 중 고등학교 야구선수의 타율과 체력 요인과의 관계에 관한 연구. **한국체육교**

- 육학회지, 7(4), 241-254.
- 오테연, 이영훈 (2016). 한국프로야구선수의 가치평가모형. **노동경제논집**, 39(2), 113-139.
- 이상학 (2023. 9. 3). 김하성과 노시환, 단 2명뿐이다... 심각한 거포 가뭄 시대에 KBO 축복 같은 '30 홈런'. **OSEN**. <https://biz.chosun.com/sports/baseball/2023/09/03/EJCGKYEIRRLCTY3PF64PXNONWM/>.
- 이장영, 강효민 (2001). 한국 프로야구 투수의 경기수행과 연봉책정의 관계. **한국스포츠사회학회지**, 14(1), 115-124.
- 장원용 (2018). 프로야구 관람결정요인의 분석: 컨조인트 분석을 중심으로. **한국엔터테인먼트산업학회 2018 추계학술대회 논문집**, 85-89.
- 장진희, 문춘걸 (2014). 한국 프로야구의 구단 승률에 대한 분석. **한국스포츠산업경영학회지**, 19(3), 17-31.
- 조선미, 김주학, 강지연, 김상균 (2023). 머신러닝(XGBoost) 기반 미국프로야구 (MLB) 의 투구별 안타 및 홈런 예측 모델 개발. **한국체육측정평가학회지**, 25(1), 65-76.
- 조형석 (2021). MLB 타자들의 스윙존에 따른 스윙선택 성향분석. 미간행 석사학위논문, 명지대학교 기록정보과학전문대학원.
- 채한승, 이종호 (2000). 프로스포츠팬 성향 경기관람결정요인에 대한 조사 연구. **한국스포츠산업경영학회지**, 5(2), 137-154.
- 최재일, 정용락 (2010). 시계열 분석에 의한 한국 프로 축구 관중수 예측(2009-2015). **한국사회체육학회지**, 39(2), 921-928.
- 한진욱, 안정찬, 오승욱, 신동일 (2015). 소셜 빅데이터를 통한 한국 프로야구 SNS 밈(meme) 분석. **한국스포츠산업경영학회지**, 20(5), 1-16.
- 홍중선, 원치환, 정동길 (2015). VUS와 HUM 최적화를 이용한 선형함수의 모수추정. **한국데이터정보과학회지**, 26(6), 1305-1315.
- Baumer, B., & Zimbalist, A. (2014). *The sabermetric revolution: Assessing the growth of analytics in baseball*. Pennsylvania, PA: University of Pennsylvania Press.
- Chang, J., & Zenilman, J. (2013). *A study of sabermetrics in major league baseball: The impact of moneyball on free agent salaries*. St. Louis, MO: Washington University in St. Louis.
- Hansen, H., & Gauthier, R. (1989). Factors affecting attendance at professional sport events. *Journal of Sport Management*, 3(1), 15-32.
- Howard, D. R., & Crompton, J. L. (2004). Tactics used by sports organizations in the United States to increase ticket sales. *Managing Leisure*, 9(2), 87-95.
- Madrigal, R. (1995). Cognitive and affective determinants of fan satisfaction with sporting events attendance. *Journal of Leisure Research*, 27(3), 205-227.
- Ng, T. W. (2017). Transformational leadership and performance outcomes: Analyses of multiple mediation pathways. *The Leadership Quarterly*, 28(3), 385-417.
- Rascher, D. A., McEvoy, C. D., Nagel, M. S., & Brown, M. T. (2007). Variable ticket pricing in major league baseball. *Journal of Sport Management*, 21(3), 407-437.
- Schulz, R., Musa, D., Staszewski, J., & Siegler, R. S. (1994). The relationship between age and major league baseball performance: implications for development. *Psychology and Aging*, 9(2), 274-286.
- Stiroh, K. J. (2007). Playing for keeps: Pay and performance in the NBA. *Economic Inquiry*, 45(1), 145-161.

ABSTRACT

A Prediction for Korean Baseball League Performance Using Intervention Time-series Analysis

Oh, Seung-wook (Kyung Hee University) · Han, Jin-wook (Kyung Hee University)

The purpose of this study was to identify trends in changes in domestic professional baseball performance and predict future performance. In order to achieve the study objective, ARIMA time series analysis was conducted using the 'R' programming language. The results are as follows. First, the prediction of changes in performance for home runs is slightly higher than the average in professional baseball leagues so far. Second, the prediction of changes in performance regarding batting average shows a much lower level than the average of domestic professional baseball leagues shown so far. Third, the prediction of changes in performance regarding RBIs is at a higher level than the average in the professional baseball league so far. Additionally discussion and practical implication were also suggested.

※ Key words : Time series analysis, ARIMA, Professional baseball, Performance, Prediction

논문 투고일 : 2023. 08. 31.

논문 심사일 : 2023. 10. 15.

게재 확정일 : 2023. 10. 31.