

# 野球における打者評価指標の経年変化と 移籍の影響

酒折 文武, 今田 一希, 山本 義郎

野球において、データに基づいて客観的な選手評価・戦術評価を行うセイバーメトリクスは一般にも広く浸透しつつある。一方で、データ計測技術や IoT の発達により、これまでよりも詳細なデータが収集されるようになり、セイバーメトリクスの世界も変化してきている。本稿では、打者に関する評価指標を中心に、近年それらがどのように進化してきたかを概説するとともに、セイバーメトリクス指標を用いた応用的な研究として、年齢による成績変化を年齢曲線として表現することを考える。そして、それを用いた移籍の効果やリーグ間の差異を記述的に考察する。

キーワード：スポーツ統計, 野球, セイバーメトリクス, 年齢曲線, 重回帰モデル, 局所重み付き回帰

## 1. はじめに

野球は比較的早い時点でデータ活用が進んだスポーツの一つである。オペレーションズ・リサーチ誌として鳩山 [1] が発表されたことは大きな意義があるが、国際的には 1970 年代にアメリカの野球ライターであった James に端を発するセイバーメトリクスが有名である。セイバーメトリクスはアメリカ野球学会の略称である SABR (Society for American Baseball Research) からの造語である。これまでは、打撃では打率・打点・本塁打数など、投球では勝利数といった選手評価に適切とは言えない指標が用いられ、いわば KKD (経験と勘と度胸) のみによって戦術選択がなされてきた。セイバーメトリクスは、この古い野球観を廃し、データに基づいてより適切な選手評価・戦術評価を行うための方法論である。そしてそれは小説や映画でも扱われ、いまでは一般にも浸透しつつある。

セイバーメトリクスでは選手の通算成績や試合のボックススコアなどの要約データを活用し、ほかの選手や球場などの影響を極力排除して、純粋に当該選手やプレイの評価を行う。たとえば、セイバーメトリクスに

おいてもっとも有名な指標の一つ OPS (On-base Plus Slugging percentages) は出塁率と長打率を加えた値であり、選手の打撃面での貢献を表すものである。チームのシーズン成績において、チーム打率やチーム出塁率と比べてチーム OPS のほうが各チームの総得点数との相関が高いことから、チーム得点増加に対する貢献を表す指標として広まり、それを個人に適用するようになった。さらには、選手の貢献を得点の予測値として捉えることができる指標として RC (Runs Created), XR (eXtrapolated Runs) などが次々と開発されていった。そして、打撃以外の投球・走塁・守備についても同じ考え方で、その選手の貢献を得失点として表すための指標が開発され、活用されてきた。

そして、近年のデータ計測技術や IoT など科学技術の発達により、セイバーメトリクスにおいても得失点の予測をより精密に行うようになってきた。ボックススコアなどよりも精密なデータが利用可能になってきたのである。たとえば、スタジアムに設置された計測器により、選手やボールの位置などの情報 (トラッキングデータ) を逐次的に自動で記録することができるようになった。アメリカのメジャーリーグ (MLB) では 2006 年以降、カメラを用いた PITCHf/x というシステムを用いて投手の投球軌道を追跡して、投手のリリースポイントや捕手の補球位置の座標、ボールの初速・終速、回転数、変化量の推定値や、ニューラルネットワークを用いた球種の判定結果などが取得される。さらに、投球のみならず HITf/x と呼ばれる打球データ、FIELDf/x と呼ばれる守備データなども取得されてきた。近年は、カメラとドップラーレーダーを併用した Statcast というシステムに変更され、投球・打撃・

さかおり ふみたけ  
中央大学理工学部  
〒 112-8551 東京都文京区春日 1-13-27  
sakaori@math.chuo-u.ac.jp  
こんだ かずき  
東海大学大学院総合理工学研究科  
〒 259-1292 神奈川県平塚市北金目 4-1-1  
8btad004@mail.u-tokai.ac.jp  
やまもと よしろう  
東海大学理学部  
〒 259-1292 神奈川県平塚市北金目 4-1-1  
yama@tokai-u.jp

走塁・守備に関する精度の高いデータが記録されている。このデータは一部、オンラインで公開されており、それによってセイバームトリクスやスポーツデータ分析がさらに活性化している。たとえば、野手の守備を評価する UZR (Ultimate Zone Rating) という指標は、各野手がどの位置でどのような強度のボールを捕球したかを用いて、同リーグの平均的な選手と比べて何失点防いだかを算出したものである。同様の考え方で打撃や走塁も、平均的な選手と比べて上積みした得点数に換算される。そして、守備位置や球場を考慮に入れたうえで、打撃・走塁・守備を合わせた得点数を勝利数に換算し、その選手が何勝分上積みしたかを表す指標 WAR (Wins Above Replacement) が計算され、野手の総合的な評価として用いられている。こうした指標については FanGraphs をはじめとするウェブ上の記事が詳しいが、近年の学術的な文献としてたとえば Marchi and Albert [2] や Albert et al. [3] がある。

一方、日本においては全く異なる状況である。ボックススコアレベルのデータこそ公開されているものの、それ以上の細かさをもったデータは一般公開されていない。一球ごとの投球の球速・座標などのログデータは人力で取得され、各種サイトで一球速報として配信されるが、それを自由に取得して分析することはできない。また近年、トラックマンというドップラーレーダーを用いた追跡システムを導入し始めたが、日本野球機構 (NPB) 主導ではなく各チームにより導入されており、導入していないチームもあるために、公開したり共同利用したりする状況ではない。このように、日本におけるセイバームトリクス研究の進展は非常に難しい状況にある。

本稿では、セイバームトリクスにおける打撃に関する評価指標の一部を概観するとともに、NPB における選手の成績の経時変化を年齢曲線として表現する。さらに、それを用いた成長タイプの分類と、リーグ移籍という外部要因による変化を考察する。野球に関する成績の年齢曲線についての研究は、たとえば Berry et al. [4] や Marchi and Albert [2] がある。

## 2. セイバームトリクスにおける主な打撃指標と打者評価方法

セイバームトリクスにおける打者の評価指標にはさまざまなものがある。ここでは、OPS, RC, XR, wOBA (Weighted On-Base Average, Tango et al. [8]) を中心に概説する。

### 2.1 OPS

たとえば、先に述べた OPS は、出塁率 (OBP, On-Base Percentage) と長打率 (SLG, Slugging percentage) との和  $OPS = OBP + SLG$  として定義される。ここで、

$$\begin{aligned} OBP &= (\text{安打} + \text{四球} + \text{死球}) \\ &\quad / (\text{打数} + \text{四球} + \text{死球} + \text{犠飛}), \\ SLG &= (\text{単打} + 2 \times \text{二塁打} \\ &\quad + 3 \times \text{三塁打} + 4 \times \text{本塁打}) / \text{打数} \end{aligned}$$

である。簡便な式の割に、チームにおける総得点と OPS との相関が比較的高いために、これを個人の指標としても用いることも多い。しかしながら、相関が高いとはいえ、各選手の貢献による得点数を直接予測する値ではない。すなわち、投球・打撃・守備・走塁を合わせて得失点への貢献を測る際には使用できない。

### 2.2 RC

得点を予測する指標としては、James [5] による RC

$$\begin{aligned} RC &= (2.4 \times C + A) \times (3 \times C + B) / (9 \times C) \\ &\quad - 0.9 \times C \end{aligned}$$

がある。ここで、

$$\begin{aligned} A &= \text{安打} + \text{四球} + \text{死球} - \text{盗塁失敗} - \text{併殺打} \\ B &= \text{塁打} + 0.24 \times (\text{四球} - \text{敬遠} + \text{死球}) \\ &\quad + 0.62 \times \text{盗塁成功} + 0.5 \times (\text{犠打} + \text{犠飛}) \\ &\quad + 0.03 \times \text{三振} \\ C &= \text{打数} \times \text{四球} \times \text{死球} \times \text{犠打} \times \text{犠飛} \end{aligned}$$

であり、それぞれ出塁数、進塁数、機会数に対応する。RC は、対象打者 1 人と、リーグの平均的な打者 8 名の計 9 名が打線を組んだ場合のチーム得点を推定し、そこから平均的打者 8 名による得点を除く形で対象打者 1 人の貢献による得点数を推定する。各係数はチームにおける各成績からチーム得点を推定する形から算出したものであり、詳しくは James [5] を参照されたい。

ここで、RC の第 1 項の分子の積をそれぞれ分母で除すと、2 項がそれぞれ OBP, SLG に近い形になっていることがわかる。すなわち、RC と OPS は、本質的には OBP と SLG の積をとるか和をとるかの違いであると見ることができる。ところが積の形であるために、これを個人の評価に当てはめようとすると、チーム成績ではみられないような個人の打撃成績を外挿する影響が強くてしまい、誤差が大きくなることが知られている。

## 2.3 XR

Furtado [6] により提案された XR は、重回帰モデルを用いてチーム得点を予測するための各変数の偏回帰係数を求めたものである。たとえば、1954～2017 年の NPB のデータ (Baseball Reference [7]) を用いると、

$$\begin{aligned} \text{XR} = & 0.43 \times \text{単打} + 0.84 \times \text{二塁打} \\ & + 1.23 \times \text{三塁打} + 1.37 \times \text{本塁打} \\ & + 0.28 \times (\text{四球} - \text{敬遠} + \text{死球}) \\ & - 0.25 \times \text{敬遠} \\ & + 0.28 \times \text{盗塁成功} - 0.16 \times \text{盗塁失敗} \\ & - 0.078 \times (\text{打数} - \text{安打} - \text{三振}) \\ & - 0.11 \times \text{三振} - 0.18 \times \text{併殺打} \\ & + 0.23 \times \text{犠飛} + 0.068 \times \text{犠打} \end{aligned} \quad (1)$$

という式を得ることができる。

XR は重回帰モデルを用いているという意味で統計学的にも理解しやすく一部で市民権を得ているが、アメリカではほとんど用いられていないようである。また、説明変数間の相関により、直感とは異なる符号の係数が得られる可能性もある。

## 2.4 wOBA

ここまでの指標は各選手のシーズン成績を用いて容易に計算できるものであったが、より詳細なデータを用いて得点の予測精度を向上させた指標が現代のセイバーメトリクスでは用いられる。その中で最も代表的なのが wOBA である。wOBA はその打者が 1 打席当たりどれだけチームの得点創出に有効な打撃をしたかを示す指標で、今日のセイバーメトリクスを扱うメディアでは打者の評価に wOBA が主だって使われている。NPB 版の wOBA として DELTA [9] では

$$\begin{aligned} \text{wOBA} = & ((0.692 \times (\text{四球} - \text{敬遠}) \\ & + 0.73 \times \text{死球} + 0.966 \times \text{失策出塁} \\ & + 0.865 \times \text{単打} + 1.334 \times \text{二塁打} \\ & + 1.725 \times \text{三塁打} + 2.065 \times \text{本塁打})) \\ & / (\text{打数} + \text{四球} - \text{敬遠} + \text{死球} + \text{犠飛}) \end{aligned}$$

と述べられている。wOBA の各係数は以下に述べるプレイの得点価値 (Run Value) に基づき計算される。

まず、プレイの得点価値について説明する。塁状況 (どの塁に走者がいるかの 8 通り)・アウトカウント (3 通り) の組み合わせによる 24 通りごとに、その状況からインニング終了までの得点数の期待値 (Run Expectancy) をデータより推定する。この得点期待値を用いて、各プレイがチーム得失点にどれだけの影響を与えたかを詳細に評価していく。たとえば、MLB でイ

ニングの先頭打者が一塁に出塁した場合、状況は「無死・無走者」(得点期待値 0.454) から「無死・一塁」(得点期待値 0.783) に変化する。よって、先頭打者のプレイにより平均得点は +0.329 増加したと評価できる。これをプレイ価値と呼ぶ。このプレイ価値を、対象期間の全プレイに対して計算し、単打・二塁打といった要素ごとに平均をとると、その要素が平均的に何得点分の価値があるかが求まる。これをプレイの得点価値という。MLB および NPB における得点期待値を表 1、表 2 に、得点価値を表 3 に記した。

各要素についての得点価値からアウトの得点価値を引いて相対的な各プレイの得点価値を算出し、それを重みとして加算した値 (仮の wOBA) の平均が、その期間の出塁率 (OBP) の平均と等しくなるように比をとって再スケーリングしたものが wOBA である。したがって、得点価値や得点期待値、出塁率がシーズンごとに異なるため、wOBA はシーズンごとに再計算する必要がある。特に、使用球やルール改正などの環境の変化があった場合には係数が大きく変化すると考えられる。

wOBA を応用した指標の一つに wRAA (Weighted Runs Above Average) がある。wRAA は、ある打者は同じ打席数のリーグ平均の打者に比べてどれだけチームの得点を増やしたかを評価する指標であり、

$$\begin{aligned} \text{wRAA} = & (\text{wOBA} - \text{リーグ平均 wOBA}) \\ & / \text{wOBA Scale} \times \text{打席} \end{aligned}$$

で定義される。ここで wOBA Scale とは、前述の比をとって再スケーリングする際の係数を意味する。たとえば、ある打者の wOBA が .353、リーグ平均が .330、その期間の wOBA Scale が 1.15 だった場合、その打者はリーグの平均的な打者よりも 1 打席平均  $(0.353 - 0.330) / 1.15 = +0.02$  チーム得点を増やしたと評価できる。すなわち、その打者に年間 500 打席を与えた場合、チーム得点を 10 得点増やすと評価できる。

## 3. 年齢曲線

選手は年齢とともに成長し、そしてピークを過ぎると徐々に衰えていくと考えられる。このことを捉えるために、打撃指標の経年変化を年齢曲線により表現することを考える。ここでは、NPB における打者の年齢曲線について考察を行う。

野球を含めた各スポーツにおける例として、Berry et al. [4] では異なる時代の選手間の比較を行うために年齢曲線を用いている。また、Marchi and Al-

表 1 1961 年度から 1977 年度の MLB における状況別の得点期待値 (Palmer [10])

	無走者	一塁	二塁	三塁	一二塁	一三塁	二三塁	満塁
無死	0.454	0.783	1.068	1.277	1.380	1.639	1.946	2.254
一死	0.249	0.478	0.699	0.897	0.888	1.088	1.371	1.546
二死	0.095	0.209	0.348	0.382	0.457	0.494	0.661	0.798

表 2 2013 年度から 2015 年度の NPB における状況別の得点期待値 (岡田 [11])

	無走者	一塁	二塁	三塁	一二塁	一三塁	二三塁	満塁
無死	0.440	0.807	1.059	1.291	1.412	1.684	1.888	2.092
一死	0.233	0.478	0.682	0.906	0.878	1.165	1.321	1.454
二死	0.087	0.204	0.305	0.349	0.417	0.495	0.578	0.758

表 3 NPB と MLB における得点価値 (岡田 [11], FanGraphs [12])

	単打	二塁打	三塁打	本塁打	四球	死球
NPB	0.437	0.786	1.117	1.408	0.292	0.311
MLB	0.70	1.00	1.27	1.65	0.55	0.57

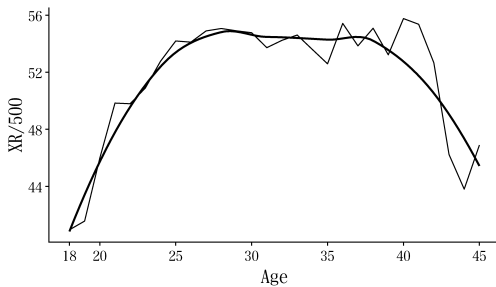


図 1 XR/500 の年齢曲線

bert [2] では、各選手の OPS の年齢曲線を二次曲線で近似し、そのピークの年齢についての簡単な考察を行っている。

### 3.1 XR の年齢曲線

前述のとおり、打撃指標として wOBA や wRAA を用いて年齢曲線を求めたいところであるが、前述の wOBA は 2013 年から 2015 年のデータに基づいた計算式であり、そのほかのシーズンでの計算式を得るためには詳細なデータが必要となる。しかしながら、データが公開されていないため、これを行うことは不可能である。そこで本稿では wOBA の代わりに XR を用いる。また、XR そのままではなく、500 打席当たりに換算した XR/500 を用いる。

図 1 は、1954 年から 2017 年のすべての打者についての年齢ごとの XR/500 の原系列と、局所重み付け回帰 (LOESS) で平滑化した年齢曲線である。なお本稿では XR の計算にはいずれのシーズンにおいても式 (1) を使用したが、シーズンごとに重回帰モデルにより XR

の式を更新することも可能である。

年齢曲線には 28 歳前後と 37 歳前後に二つのピークがあるように読み取れる。前者は選手としての実際のピークと思われる。しかし後者は、その年齢まで現役選手として活躍する選手のみが XR 算出の対象となっているという選択バイアスによるものである可能性がある。詳細はスペースの都合で割愛するが、XR ではなく打率や本塁打率などの従来指標の年齢曲線を見ると、打率は 30 代中盤に、本塁打率は 30 代後半にピークがあり、選択バイアスの問題がより一層大きく出てしまう。いずれにしても、Marchi and Albert [2] のような二次曲線での年齢曲線の表現はあまり適切ではないと思われる。

実際には、XR 自体もリーグごと、時代ごとなどの違いを考慮する必要があると思われる。図 2 はセ・リーグとパ・リーグに分けた場合の XR/500 の年齢曲線、図 3 はその年度平均の変化を見たものである。前者からセ・リーグとパ・リーグに、特に若い年代で差があることがわかる。また後者から、1980 年までの上昇傾向や、いわゆる“ラビットボール”、“違反球”と呼ばれるボールの変化などを見て取ることができる。この部分をどのように調整すべきかの課題は残るが、ここでは特段の調整を行わずに分析を進めることとする。

図 4 は、数名の有名選手の平滑化後の XR/500 の曲線である。選手によりピークや変化が異なることがわかる。たとえば、強打者タイプの王貞治・山本浩二・松中信彦は、そのピークは人により異なるものの比較的単峰型であるのに対し、巧打者タイプの長嶋茂雄・若

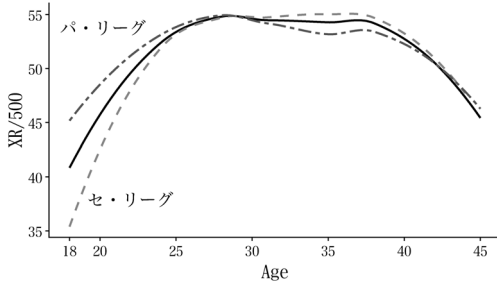


図2 リーグごとのXRの年齢曲線

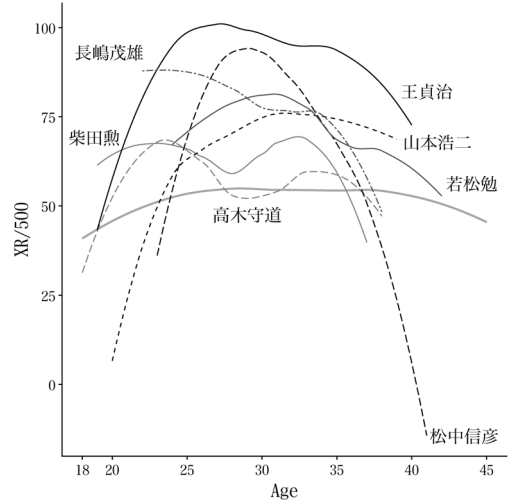


図4 タイプごとのXRの推移

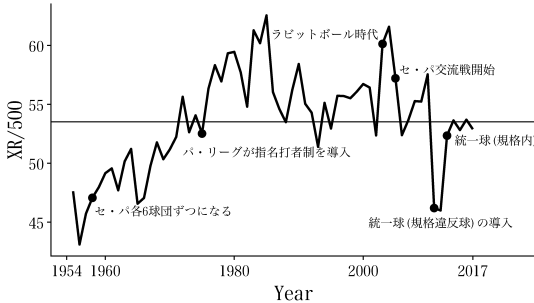


図3 平均XRの年度ごとの変化

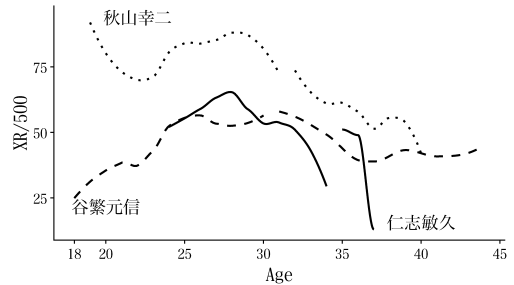


図5 リーグ内移籍した選手のXRの推移

松勉はピークから落ちた後に一度平坦になる傾向がある。さらに俊足タイプの柴田勲・高木守道は完全に二峰型である。このように、選手タイプによって、あるいは選手の特性によって、年齢曲線の型が異なると考えられる。このことをより詳しく調べるためには、年齢曲線のクラスタリングが考えられるが、選手によって現役の年齢が異なることや、不調や故障などにより1軍で出場していない時期があることには注意しなければならない。たとえば、スパースな関数データであると捉えて混合効果モデルやReduced Rank Modelなどによりモデル化する方法などが考えられる。

### 3.2 移籍による影響

さて、年齢曲線はチーム移籍という環境の変化によってどのように変わるであろうか。移籍前後の両球団で1,000打席以上出場した選手について、その前後でそれぞれ平滑化年齢曲線を求め、移籍の前後で違いがあるかどうかを見てみる。

図5は、同一リーグ内で移籍した選手の一部を示したものである。秋山幸二・谷繁元信のように移籍前と移籍直後で同程度の活躍をし、移籍後の年齢変化の傾向に変化が見られない選手が多い中、仁志敏久のように移籍して出場機会が与えられ、大きく得点貢献をする選手もいることがわかる。また、図6は、セ・リーグとパ・リーグのリーグをまたいで移籍した選手の一部を示したものである。同一リーグ内での移籍と同様

に、移籍前と同程度の活躍をする選手もいるが、和田一浩・稲葉篤紀・片岡篤史のように移籍2~3年目で大きく成績を伸ばす選手もいる。これは、リーグをまたいだ移籍により、それまでとはまったく異なる投手との対戦で成績を落とすが、徐々に慣れて本来の実力を発揮するようになったと考えられる。これらのことから、他球団から選手を補強する場合にはリーグ内から補強するのが安全であるが、初年度に思った活躍をしてくれなかった場合でも時間をかけて環境に適応していくことがわかる。

NPBとMLBの間で移籍した選手について同様に分析した結果の一部が図7である。細線がNPB、太線がMLBでの成績を表す。ここで、MLBでの値はMLBにおけるXRの式ではなく、NPBと同じものを用いた。明らかに、NPB時代と比べてMLBへ移籍後のXR/500が激減していることがわかる。これは環境の変化に加えて、NPBとMLBとのレベル差を表していると考えられる。特に井口資仁は、MLBからNPBに復帰後に再びXRが急増していることからこのこ

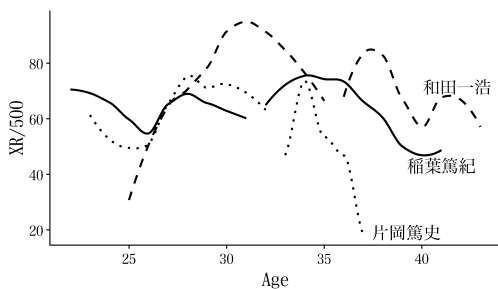


図6 リーグ間移籍した選手のXRの推移

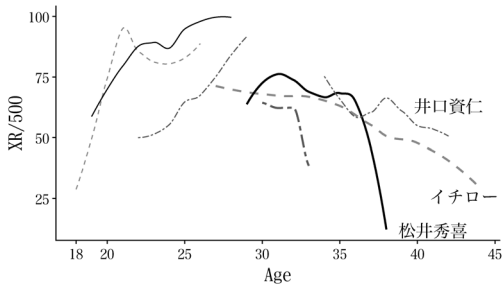


図7 MLB移籍選手のXRの年齢曲線

とがうかがえる。また、松井秀喜のように、MLB移籍後2〜3年目で大きく成績を伸ばす選手もおり、国内でのリーグ間移籍と同様に環境に適応していく様を見取ることができる。

NPBからMLBに移籍した選手13名のXR/500の変化量の平均は−20.30であり、正の値をとったのは青木宣親と田口壮のみであった。なお、青木宣親は移籍前年がいわゆる“違反球”であったためNPB最終年のXRが異常に低く、田口壮はMLB移籍初年度はAAとAAAが主でMLBは19打席しか立っていないためあまり参考にならない。一方、MLBからNPBに復帰した選手9名のXR/500の変化量の平均は20.06であり、負の値をとったのは川崎宗則と岩村明憲のみであった。なお、この2名はともに復帰後不調でNPBでの打席数が少なかった。以上を踏まえると、MLBとNPBとのXRの差は20程度と推察できる。このように、NPBとMLBの差を定量的に評価することで、MLB移籍後の選手の活躍を予測したり、助っ人外国人選手を獲得する際の成績予測を行うことが可能になると考えられる。

#### 4. おわりにかえて

本稿では、打撃に関するセイバーメトリクス指標を

概説し、その応用例として年齢曲線について述べた。また、NPBとMLB、あるいはセ・リーグとパ・リーグ間での移籍による年齢曲線の変化を記述的に見た。さらに詳しく分析するためには、前節で述べた混合効果モデルなどを用いて、移籍の効果を加えてモデリングすることが考えられる。しかしながら、現状ではパラメータの推定に十分な選手数があるとは考えられないため、上ではあくまでも記述的な分析に留めた。こうした基本的な分析に関してもまだまだ発展途上であり、さまざまな進展が考えられよう。

MLBはデータ公開が飛躍的に進んでおり、たとえばRのpitchrxパッケージやbaseballrパッケージを用いることで簡単にトラッキングデータを取得することが可能であり、さらに詳細な分析を行うことが可能である。本稿では読者の興味を考えてMLBではなくNPBの分析を例示したが、年次成績やボックススコアレベルのデータしか開示されていないという現状では限界がある。トラッキングデータを含めた詳細なデータ公開など、状況の改善が強く望まれる。

#### 参考文献

- [1] 鳩山由紀夫, “野球のOR”, オペレーションズ・リサーチ: 経営の科学, **24**(4), pp. 203–212, 1979.
- [2] M. Marchi and J. Albert, *Analyzing Baseball Data with R*, CRC Press, 2014.
- [3] J. Albert, M. E. Glickman, T. B. Swartz and R. H. König, *Handbook of Statistical Methods and Analyses in Sports*, CRC Press, 2014.
- [4] S. M. Berry, C. S. Reese and P. D. Larkey, “Bridging different eras in sports,” *Journal of the American Statistical Association*, **94**, pp. 661–676, 1999.
- [5] B. James, *The New Bill James Historical Baseball Abstract*, Free Press, 2002.
- [6] J. Furtado, “Introducing XR,” <http://www.baseballthinkfactory.org/btf/scholars/furtado/articles/IntroducingXR.htm> (2018年7月4日閲覧)
- [7] Baseball Reference, <https://www.baseball-reference.com/register/npb-stats.shtml> (2018年7月4日閲覧)
- [8] T. Tango, “Standard wOBA,” <http://www.tangotiger.com/index.php/site/article/standard-woba> (2018年7月4日閲覧)
- [9] DELTA, “wOBA,” [https://1point02.jp/op/gnav/glossary/gls\\_index\\_detail.aspx?gid=10035](https://1point02.jp/op/gnav/glossary/gls_index_detail.aspx?gid=10035) (2018年7月4日閲覧)
- [10] J. Thorn and P. Palmer, *The Hidden Game of Baseball*, Doubleday, 1985.
- [11] 岡田友輔, 「得点期待値とRun Value」, <https://1point02.jp/op/gnav/column/bs/column.aspx?cid=53003> (2018年7月4日閲覧)
- [12] FanGraphs, “Linear Weights,” <https://www.fangraphs.com/library/principles/linear-weights/> (2018年7月4日閲覧)