

# ロジスティック回帰分析を用いた打撃能力評価指標の提案

坂井 衛<sup>1</sup> 谷岡 広樹<sup>2,a)</sup> 松浦 健二<sup>2,b)</sup> 後藤田 中<sup>4,c)</sup> 和田 智仁<sup>3,d)</sup>

概要：野球では、打者の能力を判断する指標として打率の他に、本塁打を除くグラウンド内に飛んだ打球が安打になった割合である BABIP (Batting Average on Balls In Play) が使用されることがある。グラウンド内に飛んだ打球は様々な要因の影響を受けるため、ヒットかアウトの結果は投手と打者の能力のみを表しているとは考え難く、打撃結果は打者の能力のみに依存するとは言いえない。そこで本研究では、グラウンド内に飛んだ打球について、打球の方向、距離、性質、威力、打者の左右打席、走力、守備側の守備力などを説明変数としたロジスティック回帰分析を用いて、打者の能力を適切に評価できる新しい指標 OPPB(On-base plus predicted BABIP) を提案し、その有効性について評価する。

キーワード：野球，評価，ロジスティック回帰

## Proposal of Batting Ability Evaluation Index using Logistic Regression Analysis

MAMORU SAKAI<sup>1</sup> HIROKI TANIOKA<sup>2,a)</sup> KENJI MATSUURA<sup>2,b)</sup> NAKA GOTODA<sup>4,c)</sup>  
TOMOHIITO WADA<sup>3,d)</sup>

**Abstract:** In baseball, BABIP (Batting Average on Balls In Play) which is the percentage of hits that hit the ground excluding the home run is used as an index to judge the batter's ability. A hit ball is affected by various factors, so it is difficult to think that the result of a hit or out represents only the ability of the pitcher and batter. In this study, the OPPB(On-base plus predicted BABIP) based on the logistic regression model is calculated using the properties of the hit ball, the coordinates of the hit ball, the defensive strength of the opponent team, the batter's running ability, and so on.

**Keywords:** baseball, evaluation, logistic regression

### 1. はじめに

野球を観戦していると、投手が打ち取ったと思われる打球が野手の間に落ちてヒットになる場面や、逆に強い当たりの打球が野手の正面に飛んでアウトになったり、または野手のファインプレーによってアウトになるような場面をしばしば目撃する。こうした場面に着目すると、グラウ

ンド内に飛んだ打球は、打球の座標や速度、角度、野手の守備位置や守備の上手さ、風向きやグラウンドの大きさ、形など様々な要因の影響を受けており、グラウンド内に飛んだ打球の結果は、打者の能力のみに依存するとは言いえない。しかし、打率や OPS(On-base Plus Slugging) をはじめとする打者の評価指標は、いずれも打撃結果を用いており、グラウンド内に飛んだ打球が受ける様々な要因の影響を排除できていない。本研究は、グラウンド内に飛んだ打球について、打球の方向、距離、性質、威力、打者の左右打席、走力、守備側の守備力などを説明変数としたロジスティック回帰分析により、各説明変数の寄与率を求め、打者の能力を適切に評価できる新しい指標 OPPB(On-base plus predicted BABIP) を提案し、その有効性について評

<sup>1</sup> 徳島大学大学院

<sup>2</sup> 徳島大学 情報センター

<sup>3</sup> 鹿屋体育大学

<sup>4</sup> 香川大学

a) tanioka.hiroki@tokushima-u.ac.jp

b) ma2@tokushima-u.ac.jp

c) gotoda@eng.kagawa-u.ac.jp

d) wada@nifs-k.ac.jp

価する。

$$BABIP = \frac{H - HR}{AB - SO - HR + SF} \quad (1)$$

### 1.1 研究背景・目的

外的要因を考慮せず打者の成績のみを評価すると、外的要因の影響で、そのシーズンたまたま成績が良かったり、悪かったりした選手を過大評価または過小評価してしまう。首脳陣は、選手評価の源泉として成果報酬とともに、次年度以降の活躍も考慮すると考えられる。このとき、選手の能力を誤って評価してしまうと、球団の戦力構想において適切な判断ができない可能性がある。そのため本研究では、選手の能力評価のうち、最も外的要因を受けやすいと考えられる [1] 打者のグラウンド内への打球について、ロジスティック回帰分析を用いることで安定的に評価できる [2] ことを示した上で、打者の評価指標として OPPB を提案し、その他のよく用いられる打率、出塁率、長打率、OPS と比較して、より安定的に打者の評価が可能であることを示す。

## 2. 野球選手の評価指標

### 2.1 セイバートリクス

古くから野球選手の評価指標として打率、本塁打数、打点などが用いられてきたが、現在では、より客観的に選手を評価する基準・手法が提唱されている。その代表的なものがセイバートリクスである [3]。セイバートリクスとは、野球ライターで野球統計家の Bill James によって 1970 年代に提唱されたもので、野球データを統計的見地から客観的に分析し、選手の評価を行う手法である。セイバートリクスに基づく評価指標は、現在ではメジャーリーグの公式記録にも使用されている \*1。

#### 2.1.1 OPS(On-base plus slugging)

OPS とは、セイバートリクスの評価指標のうち、打者を評価する指標の一つであり、出塁率と長打率を足しあわせたものである [4]。OPS は、比較的簡単に求められる指標でありながらも、得点と OPS の相関係数が、得点と打率の相関係数よりも高いことから、打者の評価指標として普及し、現在ではメジャーリーグでは公式記録に採用されている。しかし、OPS は実際の出塁率と長打率を足し合わせたものであり、実際の打撃結果に基づく指標のため、グラウンド内に飛んだ打球が外的要因を受けた場合を考慮されていない。

#### 2.1.2 BABIP(Batting Average On Ball In Play)

グラウンド内に飛んだ打球についての評価指標としては、1990 年代に Voros McCracken により BABIP という評価指標 [1] が提唱された。BABIP は式 (1) で表され、本塁打を除くグラウンド内に飛んだ打球が安打になった割合を示す。

ここで  $H$  は安打数、 $HR$  は本塁打数、 $AB$  は打席数、 $SO$  は三振数、 $SF$  は犠飛数である。Voros McCracken の研究 [1] では、投手の成績を守備に依存しないものと依存するものに分け、BABIP は投手の評価にとって意味を持たないことを示すための分析を行った。結果として守備に依存しない成績、与四球や奪三振、被本塁打といったものについては年度ごとに一貫性が見られたが、守備に依存する項目である BABIP では一貫性がなくランダムであった。つまり選手の能力は BABIP に強く影響しないとしている。同様に打者についても四球や三振、本塁打に関しては年度ごとの相関が高いのに対し BABIP は相関が低いことが報告されている [5] [6]。

このことから、グラウンド内に飛んだ打球がヒットになるかアウトになるかは選手の能力のみでは定まらず、外的要因による影響を受けている。そのため、グラウンド内への打球を多く放つ選手の能力を適切に評価するためには、グラウンド内の状況を考慮した指標が必要である。

### 2.2 関連研究

Chris Dutton らの研究 [7] では、線形回帰分析を用いて選手の BABIP の予測を試みている。予測した BABIP より実際の BABIP の方が高かった選手を幸運な選手、実際の BABIP の方が低かった選手を不幸な選手として、それぞれ選手の次年度の BABIP の変化を分析している。結果として、幸運な選手は次年度で BABIP が下降し、不幸な選手は次年度で BABIP が上昇することが判明した。

この研究の問題点として、説明変数に打球の座標や野手の守備力といったグラウンド内の情報が使われておらず、最終的な打撃結果の組み合わせにすぎないこと、選手の走力を表す指標にスピードスコアを用いているが、これは盗塁数や得点数を用いて簡易的に選手の走力を示すものであり、打球をヒットにするための走力、ホームベースから一塁までの走力を正確に表してるとはいえないことが挙げられる。

実際、不幸な選手の中の一人にイチローを挙げている。イチローは本塁打、三振、四死球が少なく、内野安打が非常に多い選手のため BABIP も高い値を示しているが、グラウンド内の外的要因のために、BABIP では打撃能力を安定的に評価できないといえる。

## 3. 提案手法

### 3.1 分析の流れ

本研究では、打席データを学習データとして用い、その打席結果を目的変数としてロジスティック回帰分析を用いて回帰モデルを作成する。作成した回帰モデルから、観測

\*1 "MLB.com The Official Site of Major League Baseball"  
<https://www.mlb.com/>

可能な情報に基づき各打席の安打確率を求め、選手の予測BABIPを算出する。次に、各選手の予測BABIPと実際のBABIPを比較することで、過大評価・過小評価している選手を抽出する。さらに、次年度の選手のBABIPを予測することで選手の打撃能力を安定的に評価できることを示す。その上で新たな評価指標としてOPPBBを提案し、既存指標と比較して、より安定的に打者の評価が可能であることを示す。

### 3.2 ロジスティック回帰分析

グラウンド内の状況を考慮した各打席の安打確率を求めるために、回帰モデル作成にロジスティック回帰分析を用いる。ロジスティック回帰分析の説明変数には、量的変数、質的変数のどちらも用いることができ、打席データに対し、グラウンド内の状況を説明変数、その打席がヒットであったかどうか(1または0)を目的変数とする。

ロジスティック回帰分析の出力値は0から1の値をとるため、これを打球がヒットになる確率とみなせる。ロジスティック回帰モデル式(2)は以下のように表される。

$$\log \frac{p}{1-p} = b_0 + \sum_{j=1}^k b_j x_j \quad (2)$$

これを  $p$  について変形すると以下のように表される。

$$p(x) = \frac{1}{1 + \exp\left(-b_0 - \sum_{j=1}^k b_j x_j\right)} \quad (3)$$

ここで、 $p(x)$  は  $x$  に基づく予測値、 $x = x_1, \dots, x_k$  は説明変数、 $b_0$  は定数、 $b_1, \dots, b_k$  は回帰係数である。

### 3.3 予測BABIP

ある選手の全打席結果の安打予測値の平均を予測BABIPと定義する。予測BABIPは式(4)で表され、これはBABIPの内野に飛んだ打球の打撃結果を、回帰モデルによる予測値に置き換えたものである。

$$\text{Predicted BABIP} = \frac{\sum p}{AB - SO - HR + SF} \quad (4)$$

ここで  $\sum p$  は予測値の総和、 $AB$  は打席数、 $SO$  は三振数、 $HR$  は本塁打数、 $SF$  は犠飛数である。

### 3.4 OPPB(On-base plus predicted BABIP)

回帰モデルを用いて算出した予測値を用いて新たな打者の評価指標を提案する。既存の評価指標に含まれる、外的要因の影響を受けている打撃結果の部分を予測値に置き換えることで、より安定した評価指標を作成する。OPPBBの式(5)で表され、これは出塁率に含まれるグラウンド内に飛んだ打球の結果を予測値に置き換えたものである。

$$\text{OPPBB} = \frac{\sum p + BB + HR}{AB + SO + HR + BB} \quad (5)$$

ここで  $\sum p$  は予測値の総和、 $BB$  は四死球数、 $HR$  は本塁打数、 $AB$  は打席数、 $SO$  は三振数である。

## 4. 実験

### 4.1 使用データ

使用するデータは日本統計学会、及びスポーツ統計分科会が主催する第9回スポーツデータ解析コンペティション<sup>\*2</sup>に参加することでデータスタジアム株式会社<sup>\*3</sup>から提供されるものであり、2016, 2017, 2018年の日本プロ野球の公式戦2574試合のデータである。グラウンド内に飛んだ打球のみを分析するために、打席データから「打撃区分名」が、安打、長打、凡打であるものに限定する。

### 4.2 説明変数

説明変数には、打者がボールを打った後に打球がヒットかアウトになるかを左右する情報を用いる。本研究では、データスタジアム株式会社から提供された打席データの中から、ヒットかアウトになるかを左右する要因と考えられるものを用い、現時点で観測可能な限りの要因を考慮した予測モデルを作成する。

#### 4.2.1 打球の座標

ゴロ、フライ、ライナーごとの座標  $(x, y)$  を極座標  $(r, \theta)$  に変換したものを使用する。ホームベースの座標は  $(x, y) = (0, 0)$  とする。また座標は打球の第一捕球者の捕球位置を示す。x座標 y座標ともに  $0 \leq x \leq 240$ ,  $0 \leq y \leq 240$  の範囲をとり、 $r$  は  $0 \leq r \leq 240$ ,  $\theta$  は  $0 \leq \theta \leq \pi$  の範囲をとる。

#### 4.2.2 打球の強さ

同じ座標に飛んだ打球であっても、打球に強さによって打撃結果は変わる可能性があるため、説明変数として定義する。提供データの中から「打球強度」の項目を使用する。各打球の強さを無、弱、中、強の四段階に分けたものである。元データはテキストデータであるので、1から4の数値に置き換えた。

#### 4.2.3 各塁のランナー状況

ランナーがない場合、通常ヒットになる打球であったとしても、塁上のランナーがアウトになれば打撃結果は凡打になってしまうため、ランナー状況はヒットかアウトかを左右する要因と考えられる。各塁にランナーがいれば1、いなければ0とする。

#### 4.2.4 対戦相手チームの守備力

打球の処理に関わった野手の個々の守備力を算出することが困難だったため、守備力にはチーム全体の守備力の指標であるDER (Defense Efficiency Rating) [8]を用いた。

<sup>\*2</sup> "日本統計学会 スポーツ統計分科会" estat.sci.kagoshima-u.ac

<sup>\*3</sup> "データスタジアム株式会社" www.datastadium.co.jp

次式が DER の算出方法である .

$$DER = \frac{AB - H - BB - SO - E}{AB - HR - BB - SO} \quad (6)$$

ここで  $AB$  は打席数,  $H$  は安打数,  $BB$  は四死球数,  $SO$  は三振数,  $HR$  は本塁打数,  $E$  は失策数である .

表 1 と表 2 が 2016, 2017 年度それぞれの 12 球団の平均 DER を算出したものである .

#### 4.2.5 打者の走力

ボールを打った後, 打者がヒットかアウトかをコントロールできる唯一の要因として, 走力が考えられる . もし, ほとんどのケースでアウトになっている場所にボールを打ったとしても, 選手の足が速ければ内野安打になる場合があるため, 打者の走力を説明変数に用いる . 打者の走力を示す数値として, 打者がボールを打ってから一塁を踏むまでの一塁到達時間を使用する . 実際の試合の映像から選手が全力疾走したとみられる打席の一塁到達時間を計測した . 試合の映像は, 2018 年のプロ野球の試合 108 試合分であり, その中で計測可能であった 92 人の選手の一塁到達時間を用いる .

#### 4.2.6 左右打席

左打席と右打席では一塁までの距離が異なるため, 安打を左右する要因と考えられる . 左打席なら 1, 右打席なら 0 とする .

#### 4.3 評価方法

実際の BABIP と予測 BABIP の差をみることで, 選手を評価する . 予測 BABIP より実際の BABIP が高ければ過大評価されている選手, 実際の BABIP が低ければ過小評価されている選手とする .

#### 4.4 回帰モデル作成

学習に使用したデータは打席データから「打撃区分名」が, 安打, 長打, 凡打であるものであり, 一塁到達時間を計測できた選手 92 人の選手の 2016, 2017 年度の打席, 合計で 47155 打席のデータである . 得られた説明変数の係数を式 (2) に当てはめた回帰モデル式は以下 (7) のようになる . 表 3 は, R による回帰分析の出力結果であり, 推測された係数及び関連の統計量を示している .

$$\log \frac{p}{1-p} = -0.978 + 0.0362x_1 + 0.0486x_2 + \dots - 6.289x_{12} - 0.491x_{13} \quad (7)$$

表 3 の数値は, 左から各変数の係数推定値, 標準誤差,  $z$  値,  $p$  値を示す . 係数は, 正であれば変数が増加するほど確率を上げ, 負であれば確率を下げる方向に働く .  $p$  値は係数の出現確率を示し,  $z$  値と  $p$  値によって各説明変数とモデルの統計的有意性を示すことができる . 左右打席を除く, 全ての回帰係数に関して,  $p < 0.05$  であるため, これ

表 3 回帰分析結果

	Estimate	std.Error	z value	p value
定数	-0.978	0.8711	-1.123	0.2616
左右打席	0.0362	0.0358	1.014	0.3107
ゴロ $r$	0.0486	0.0006	83.042	2E-16
ゴロ $\theta$	0.6713	0.0500	13.430	2E-16
フライ $r$	0.0321	0.0004	73.731	2E-16
フライ $\theta$	0.3884	0.0427	9.079	2E-16
ライナー $r$	0.0619	0.0010	61.758	2E-16
ライナー $\theta$	-0.5669	0.1305	-4.343	1.40E-05
打球強度	0.1505	0.0270	5.573	2.50E-08
一塁	0.0701	0.0310	2.263	0.0236
二塁	0.0769	0.0354	2.171	0.0299
三塁	0.3227	0.0486	6.633	3.28E-11
相手チーム DER	-6.289	1.165	-5.400	6.68E-08
到達時間	-0.491	0.0811	-6.054	1.42E-09

らの説明変数は統計的に有意である . 左右打席の係数に有意ではない理由として, 左打席に立つことで一塁との距離が短くなりヒットになりやすいという情報が, 到達時間によって説明できるため, 統計的に重要でない変数とみなされたと考えられる . 実際, 到達時間を説明変数から除いて回帰モデルを作成すると, 左右打席は有意な変数とみなされた . 各塁ランナー状況に関して, 3 つとも 2 値データであるが塁が進むほど係数が増加している傾向が見られる . これは三塁にランナーがいる状況が他と比べて少ないことや, 犠牲フライや前進守備などでアウトになりにくい状況が作られやすいことが影響していると考えられる .

#### 4.5 BABIP と予測 BABIP の比較

回帰分析によって求めたパラメータに基づく回帰モデルの式を用いて, 各打席の予測値を算出した . この予測値を基に, 各選手の予測 BABIP を算出する . 分析対象の打者は, 年間の打数が 200 打数以上の選手 (2016 年 57 人, 2017 年 66 人) である . 表 4 から表 7 は, 各年度ごとに, 予測 BABIP よりも実際の BABIP が良かった選手, 予測 BABIP よりも実際の BABIP が悪かった選手の上位 10 名を列挙したものである . つまり各年度で過大評価している選手, 過小評価している選手である .

### 5. 評価

本研究で作成した回帰モデルによる予測 BABIP が選手に適正な評価を与えているかどうかについて検証する .

#### 5.1 次年度の成績予測

2016 年度で過大評価と判定された選手と過小評価と判定された選手の 2 群に分け, 2 年間での BABIP と予測 BABIP の変化を調べることで, 選手の次年度の成績を予測する . グラウンド内の要因による影響で自身の能力以上, もしくはそれ以下の成績になってしまっているため,

表 1 2016 年度 DER

巨人	阪神	中日	広島	DeNA	ヤクルト	ソフトバンク	西武	日本ハム	オリックス	ロッテ	楽天
0.68	0.687	0.696	0.7	0.688	0.686	0.707	0.674	0.703	0.678	0.685	0.668

表 2 2017 年度 DER

巨人	阪神	中日	広島	DeNA	ヤクルト	ソフトバンク	西武	日本ハム	オリックス	ロッテ	楽天
0.694	0.677	0.699	0.696	0.692	0.683	0.72	0.701	0.692	0.692	0.682	0.697

表 4 2016 年度 過大評価されている選手

	名前	BABIP	予測 BABIP	差
1	大引 啓次 (ヤ)	.319	.269	.05
2	高山 俊 (阪)	.346	.298	.048
3	原口 文仁 (阪)	.331	.285	.046
4	田村 龍弘 (口)	.336	.291	.045
5	菊池 涼介 (広)	.378	.336	.042
6	陽 岱鋼 (日)	.368	.331	.037
7	今江 敏晃 (楽)	.308	.271	.037
8	坂口 智隆 (ヤ)	.342	.307	.035
9	大島 洋平 (中)	.328	.295	.033
10	中島 卓也 (日)	.331	.301	.03

表 6 2017 年度 過大評価されている選手

	名前	BABIP	予測 BABIP	差
1	藤井 亮太 (ヤ)	.316	.249	.067
2	京田 陽太 (中)	.324	.272	.052
3	田中 広輔 (広)	.364	.32	.044
4	高山 俊 (阪)	.315	.273	.042
5	雄平 (ヤ)	.356	.314	.042
6	亀澤 恭平 (中)	.321	.283	.038
7	大田 泰示 (日)	.297	.265	.032
8	茂木 英五郎 (楽)	.344	.313	.031
9	小谷野 栄一 (オ)	.309	.279	.03
10	坂口 智隆 (ヤ)	.336	.307	.029

表 5 2016 年度 過小評価されている選手

	名前	BABIP	予測 BABIP	差
1	岡島 豪郎 (楽)	.279	.355	-.076
2	清田 育宏 (口)	.27	.311	-.041
3	中村 悠平 (ヤ)	.23	.252	-.039
4	近藤 健介 (日)	.316	.348	-.032
5	鳥谷 敬 (阪)	.277	.306	-.029
6	西野 真弘 (オ)	.294	.322	-.028
7	ビシエド (中)	.284	.309	-.025
8	丸 佳浩 (広)	.334	.355	-.021
9	鈴木 大地 (口)	.316	.332	-.016
10	中村 晃 (ソ)	.314	.33	-.016

表 7 2017 年度 過小評価されている選手

	名前	BABIP	予測 BABIP	差
1	栗山 巧 (日)	.279	.355	-.076
2	岡島 豪郎 (楽)	.27	.311	-.041
3	福留 孝介 (阪)	.23	.252	-.039
4	西野 真弘 (オ)	.316	.348	-.032
5	丸 佳浩 (広)	.277	.306	-.029
6	角中 勝也 (口)	.294	.322	-.028
7	ビシエド (中)	.284	.309	-.025
8	安達 了一 (オ)	.334	.355	-.021
9	藤田 一也 (楽)	.316	.332	-.016
10	嶋 基宏 (楽)	.314	.33	-.016

過大評価されている選手であれば次年度成績が下がり、過小評価されている選手であれば次年度成績が上がると予想できる。

### 5.1.1 2群に分ける基準

2016, 2017 年ともに 200 打数以上あった選手計 46 名に対し、過大評価と判定された選手 20 人、過小評価と判定された選手 26 人の 2 群に分ける。

### 5.2 過大評価選手の分析

過大評価と判定された選手 20 人の 2016, 2017 年の BABIP に対し、対応のある 2 標本の t 検定を行った。帰無仮説は「2016 年と 2017 年の BABIP の平均に差はない」とする。対立仮説は「2017 年の BABIP は 2016 年の BABIP よりも低い」である。有意水準は 0.05 とする。検定を行った結果の各検定統計量は、表 8 に示したとおりとなった。結果として、 $p = 0.00006 (< 0.05)$  となり、帰無仮説を棄却し、2017 年に BABIP が減少したといえる。また効果量は 1.06 と大きな効果を示した。

### 5.3 過小評価選手の分析

過小評価と判定された選手 26 人の 2016, 2017 年の BABIP に対し、対応のある 2 標本の t 検定を行った。帰無仮説は「2016 年と 2017 年の BABIP の平均に差はない」とする。対立仮説は「2017 年の BABIP は 2016 年の BABIP よりも高い」である。有意水準は 0.05 とする。検定を行った結果の各検定統計量は、表 11 に示したとおりとなった。結果として、 $p = 0.159 (\geq 0.05)$  となり、帰無仮説を棄却できず、BABIP が上昇したとはいえない。また効果量は 0.199 と小さな効果を示した。

2017, 2018 年についても同様な検定を行った。結果として、過大評価している選手の BABIP は減少し、過小評価している選手の BABIP は上昇し、2016, 2017 年と同様な傾向がみられた。

### 5.4 連続する 2 年間の相関

2016, 2017 年ともに 200 打数以上あった選手計 46 名、及び 2017, 2018 年ともに 200 打数以上あった選手計 52 名

表 8 過大評価選手の BABIP 検定結果

	2016 年 BABIP	2017 年 BABIP
平均	0.345	0.313
分散	0.0008	0.0010
t 値	4.764	
P 値 片側	0.00006	
t 境界値 片側	1.729	
P 値 両側	0.0001	
t 境界値 両側	2.093	
効果量 d	1.065	

表 11 過小評価選手の BABIP 検定結果

	2016 年 BABIP	2017 年 BABIP
平均	0.304	0.310
分散	0.0010	0.0012
t 値	-1.017	
P 値 片側	0.159	
t 境界値 片側	1.708	
P 値 両側	0.319	
t 境界値 両側	2.059	
効果量 d	0.199	

に対して t, BABIP 及び予測 BABIP の相関関係を調べた。表 9 は各年度間での BABIP と予測 BABIP の相関係数を示している。どの年度も BABIP の相関係数に比べ、予測 BABIP の相関係数が高いことから、予測 BABIP は選手本来の能力を表し、次年度での成績を予測できる可能性があるといえる。

表 9 連続する 2 年間での相関係数

2016-2017 BABIP	0.45
2016-2017 予測 BABIP	0.73
2017-2018 BABIP	0.45
2017-2018 予測 BABIP	0.67

## 5.5 既存指標との比較

予測 BABIP を用いることで、実際の打撃結果を基に算出された指標と比較して安定的に選手を評価できることが示された。その上で、既存の評価指標と比較してより安定的に評価ができる指標を目指す。表 10 は各年度間での OPS と OPPB の相関係数を示している。どの年度も OPS の相関係数はと予測 BABIP よりも強い相関があるが、OPPB の相関係数は OPS を上回る強い相関を示している。

表 10 連続する 2 年間での相関係数

2016-2017 OPS	0.76
2016-2017 OPPB	0.83
2017-2018 OPS	0.72
2017-2018 OPPB	0.77

## 6. まとめ

本研究では、グラウンド内の情報を説明変数、打撃結果を目的変数としてロジスティック回帰分析することで、外的要因を考慮した予測 BABIP を作成した。予測 BABIP と実際の BABIP を比較することで、過大評価している選手、過小評価している選手を抽出した。また、予測 BABIP は実際の BABIP に比べ、年度間の相関が強く、安定しており、次年度の成績を予測することができる。さらに、予測 BABIP を用いた新指標である OPPB は、OPS よりも高い相関を示し、打者をより安定的に評価できる指標であ

るといえる。

謝辞 この研究はデータスタジアム株式会社、情報・システム情報機構統計数理研究所から支援を受けたものです。本研究は JSPS 科研費 JP18H03344 の助成を受けたものです。

## 参考文献

- [1] Voros McCracken: *Pitching and Defense: How Much Control Do Hurlers Have?*, 入手先 <https://www.baseballprospectus.com/news/article/878/pitching-and-defense-how-much-control-do-hurlers-have/>, (参照 2018-02-14)
- [2] Sakai Mamoru, Hiroki Tanioka, Kenji Matsuura, Masahiko Sano, Kenji Ohira, Tetsushi Ueta and Sakaguchi Hiroaki : *Evaluating Hitting Skills of NPB Players with Logistic Regression Analysis*, CSCE'18 Proceedings, 313-319, Las Vegas, Jul. 2018.
- [3] Richard J. Puerzer: *From Scientific Baseball to Sabermetrics: Professional Baseball as a Reflection of Engineering and Management in Society*, INE: A Journal of Baseball History and Culture, 2002.
- [4] *What is a On-base Plus Slugging (OPS)?*, 入手先 <http://m.mlb.com/glossary/standard-stats/on-base-plus-slugging>, (参照 2018-02-14)
- [5] Hiroya Sasaki: *The meaning of BABIP and the Difficulty of the Interpretation*, 入手先 <http://www.baseball-lab.jp/column/entry/175/>, (参照 2018-02-14)
- [6] Matt Klaassen: *Basic Hitting Metric Correlation 1955-2012, 2002-2012*, 入手先 <https://blogs.fangraphs.com/basic-hitting-metric-correlation-1955-2012-2002-2012/>, (参照 2020-01-25)
- [7] Chris Dutton: *Batters and BABIP*, 入手先 <https://www.fangraphs.com/tht/batters-and-babip/>, (参照 2018-02-14)
- [8] *Defensive Efficiency Ratio (DER)*, 入手先 <http://m.mlb.com/glossary/advanced-stats/defensive-efficiency-ratio/>, (参照 2018-02-14)