OPSの計算式における 出塁率と長打率の最適な価値配分 ~一見手抜きだが便利な指標~

芝浦工業大学 数理科学研究会 BV17057 西脇 友哉 平成 30 年 5 月 19 日

目 次

1	導力	導入,用語解說										
	1.1	前提	1									
	1.2	公式記録	1									
		1.2.1 出塁率 (OBP) On-base percentage	1									
		1.2.2 長打率 (SLG) Slugging percentage	2									
	1.3	セイバーメトリクス指標	2									
		1.3.1 OPS (On-base Plus Slugging)	2									
		1.3.2 NOI (New Offensive Index)	3									
		1.3.3 IsoP (Isolated Power)	3									
2	主成分分析による出塁率と長打率の総合化											
	2.1	基本方針	3									
	2.2	対象データについて	3									
	2.3	主成分分析の手順	3									
		2.3.1 第1主成分	5									
		2.3.2 第 2 主成分	5									
3	主成分得点の持つ意味											
	3.1	因子負荷量による判断	5									
	3.2	指標と得点の相関・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	6									
4	まと	b b	6									
5	補足	·	7									

研究背景

昨今のプロ野球においてセイバーメトリクスが浸透しつつあるが、数ある指標のうち特に **OPS** は、その簡便性 ゆえ評論家のみならずネットユーザーにも広く認知され、打率や本塁打数ではなく **OPS** の数値を目標に掲げる選手まで登場する程に定着した。 **OPS** の計算式は「(**OPS** =) 出塁率 (**OBP**) + 長打率 (**SLG**)」であるが、「出塁率と長打率を等しい重み付けで総合化する」というこの原理は「複数の変量を総合して主成分を求める」という主成分分析の発想に近いものがあるのではないかと考え、研究の着手に至った。

1 導入,用語解説

1.1 前提

• 余白の都合により、球団名の略称の表記を以下の通りに使用する.

	パシフィック・リーグ	セントラル・リーグ			
ソ	福岡ソフトバンクホークス	広	広島東洋カープ		
西	埼玉西武ライオンズ	神	阪神タイガース		
楽	東北楽天ゴールデンイーグルス	De	横浜 DeNA ベイスターズ		
オ	オリックス・バファローズ	巨	読売ジャイアンツ		
日	北海道日本ハムファイターズ	中	中日ドラゴンズ		
口	千葉ロッテマリーンズ	ヤ	東京ヤクルトスワローズ		

- 野球の指標のうち、打率、守備率、勝率など、割合を表すものを表記する際には1の位の「0」を省略することが一般的である。(例:打率2割8分6厘→.286)
 本研究において、過去の選手の成績を提示する際には慣例に従い「0」を省略して表記しているが、計算式中において指標を用いる際には、数値としての側面を意識し、省略せずに表記している。
- 過去の試合結果をはじめ研究に必要なデータについては NPB (日本野球機構) の公式 HP から収集した.

1.2 公式記録

打率, 打点, 本塁打, 四球, 犠打などが打撃の公式記録に該当する. これらは従来から打者の評価に用いられてきた指標であり, 特に打率, 打点, 本塁打を総称して「打撃3部門」とよばれるなど, 長いプロ野球の歴史において確固たる地位を築いてきた. ここでは OPS の導出に用いられる出塁率と長打率の解説を行う. なお, 「最高出塁率」というタイトルは存在するが. 長打率に関するタイトルは存在しない.

1.2.1 出塁率 (OBP) On-base percentage

計算式 1 (出塁率).

出塁率 = $(安打 + 四球 + 死球) \div (打数 + 四球 + 死球 + 犠飛)$

出塁率は文字通り「出塁する確率」を表すが、「アウトにならない確率」と解釈をすることもできる. 野球で点を取る手段は「適時打」「本塁打」「犠飛」など複数存在するが、全てに共通するのは「3つ目のアウトを取られる前に走者を本塁へ返さなければならない」ということである. 故に攻撃においてアウトにならないことは極めて重要であり、そこに四死球の価値が見出される. かつて出塁率は打率に比べて注目度が低かったが、近年では打撃指標としての有用性が見直されている.

なお、 犠飛は打数に含まれないため打率の計算では分母から除外されているが、 出塁率の計算においては犠飛は 分母に含まれるため注意が必要である. 四球と犠飛の数によっては出塁率が打率よりも低くなる例が存在しうる. **例 1.1.** '00 の荒木雅博選手(中)が打率 .200, 出塁率 .167 を記録している. (打席数 12, 打数 10, 安打数 2, 四死 球 0, 犠飛 2)

1.2.2 長打率 (SLG) Slugging percentage

計算式 2 (長打率).

長打率 = 塁打÷打数

計算式 3 (塁打).

長打率は「長打を打つ確率」ではなく、計算式上の意味合いとしては「1 打数につき稼いだ進塁の数」を表しているため、長打率が 1.000 を超えない限りは内野安打でも長打率は上昇する。また、長打率が最大となるのは全打数で本塁打を打った場合なので最大値は 4.000 だが、NPB 歴代最高は .779('13 W. バレンティン選手 (ヤ))である。打率では安打はすべて等しい価値として計算されるが、長打率の計算においては単打、二塁打、三塁打、本塁打にそれぞれ異なる重み付けがなされるため、出塁の頻度が低くても塁を稼げる(=長打力のある打者)を評価することができる。

例 1.2. '17 に鳥谷敬選手(神) は出塁率 .390, 長打率 .377 であったのに対し, 鈴木誠也選手(広) は出塁率 .389, 長打率 .547 であった. この 2 選手の成績を比較すると, 出塁率はほぼ同じであるが, 長打率は鈴木選手が鳥谷選手を大きく上回っている. この事から, 鈴木選手は鳥谷選手よりも「塁を稼ぐ効率が良い」といえる.

1.3 セイバーメトリクス指標

セイバーメトリクス (SABR metrics) は、アメリカ野球学会の略称である「SABR」と測定法を意味する「metrics」を足した造語であり、1980 年にビル・ジェームズ氏により提唱された統計学的見地による野球の分析手法である。選手の成績から得点と失点への貢献度、ひいてはその選手に何勝分の価値があるかを計ることを主とし、その分析のために様々な指標を用いる。そして本研究の主題である「OPS」は打撃成績を評価する指標の1つである。

セイバーメトリクス指標は「打撃」「投球」「守備」「走塁」それぞれについて非常に多岐に渡って存在するものであり、この章で解説する指標はそれらの内のほんの一部に過ぎないという点は注意されたい.

1.3.1 OPS (On-base Plus Slugging)

計算式 4 (OPS).

OPS = 出塁率 + 長打率

OPS は、前述の出塁率と長打率を同時に評価することができ、いわゆる「強打者」が高い数値を示す指標である. 理論上の最大値は 5.000(出塁率: 1.000, 長打率: 4.000)だが、NPB、MLB ともに例年リーグの平均は .700 程度である. OPS が提唱された背景には「打撃 3 部門」の指標としての欠陥があり、特に打率に代わる評価基準として、選手の年俸の査定や獲得選手の選考などに用いられるほどに定着した.

出塁率と長打率をそのまま足しただけという非常にシンプルな構造でありながら得点との連動性の高さに非常に優れている点が他の指標にはない長所である。データスタジアム株式会社『野球×統計は最強のバッテリーである セイバーメトリクスとトラッキングの世界』によると、NPB の過去 30 年分のシーズン成績によるチーム OPS と 1 試合の平均得点との重相関係数は $R^2=0.901$ であった。それに対し出塁率と平均得点では $R^2=0.7284$ 、長打率と平均得点では $R^2=0.8182$ であり、OPS による評価の精度の高さがうかがえる。

1.3.2 NOI (New Offensive Index)

計算式 5 (NOI).

$$NOI = (出塁率 + 長打率 \div 3) \times 1000$$

一部には「OPS は出塁率を過小評価している」という意見が存在し、NOI は OPS の見直し作業から考案された。Wikipedia によると、考案者であるポール・デポデスタ氏は、"過去のメジャーでのチームデータを基に、「出塁率と長打率の割合を 3:1 にすれば、両者の和である NOI は OPS よりも実際のチームの総得点数との近似性がさらに高まる」という推論を導き出した。"とされている。

1.3.3 IsoP (Isolated Power)

計算式 6 (IsoP).

$$IsoP = 長打率 - 打率$$
$$= (二塁打 + 三塁打 \times 2 + 本塁打 \times 3) \div 打数$$

公式記録の長打率は重要度の高い指標であるが、その名の通りに純粋な長打力を表す訳ではない. IsoP は長打率から単打の要素を排除することで長打力の評価を可能にしている.

2 主成分分析による出塁率と長打率の総合化

2.1 基本方針

OPS の仕組みからヒントを得て、主成分分析により「出塁率と長打率を総合して評価する」ことを主目的とする。本研究では得られた主成分得点を指標の一つとして扱う事とし、「OPS」「NOI」「第1主成分得点」「第2主成分得点」の4指標それぞれについてチームの1試合平均得点との相関係数を計算し、どの指標が優れているかについて考察する。

2.2 対象データについて

前章では OPS を打者個人の成績として紹介したが、個人の成績を対象とする場合得点との相関関係を調べることが難しくなるため、チームの 1 試合平均得点を採用することとした. 1

具体的には、NPB の公式 HP に掲載されている' $12\sim'17$ のペナントレースにおける各球団の打撃成績から「チームの出塁率」「チームの長打率」を引用し、分析に用いる。 すなわち、データの個数は $12(球団)\times 6(年)=72(個)$ となる。 詳細なデータについては、末尾の補足資料を参照されたい。

2.3 主成分分析の手順

以下, x_1 は出塁率, x_2 は長打率を表すものとする.

主成分とは、サンプルの違いを最もよく表現できるように説明変数を合成したものである. ここで、説明変数の 線形関数

$$p = w_1 x_1 + w_2 x_2$$

を考え、主成分得点pの分散 V_p が最大になるような係数 w_1, w_2 を求める. ただし、

$$w_1^2 + w_2^2 = 1$$

という条件が生じるものとする.

¹一応打撃・走塁成績による得点への貢献度を計る「RC (Runs Created)」という指標が存在するため不可能ではないが、RC は計算式が極めて複雑であり、その計算に労力をかけることが、主目的である OPS の考察に支障をきたしかねないという結論に至った.

公式 1 (分散の公式).

$$Var(ax + by) = a^{2}Var(x) + b^{2}Var(y) + 2abCov(x, y) \quad (a, b : const.)$$

証明.

$$Var(ax + by) = \frac{1}{n} \sum_{i=1}^{n} \left(ax_i + by_i - (\overline{ax + by}) \right)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(ax_i + by_i - (\overline{ax} + b\overline{y}) \right)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(a(x_i - \overline{x}) + b(y_i - \overline{y}) \right)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left\{ a^2(x_i - \overline{x})^2 + b^2(y_i - \overline{y})^2 + 2ab(x_i - \overline{x})(y_i - \overline{y}) \right\}$$

$$= a^2 \cdot \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 + b^2 \cdot \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^2 + 2ab \cdot \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

$$= a^2 \text{Var}(x) + 2ab \text{Cov}(x, y) + b^2 \text{Var}(y)$$

この公式を利用して分散 V_p を式で表すと,

$$V_p = \text{Var}(w_1 x_1 + w_2 x_2)$$

= $w_1^2 \cdot \text{Var}(x_1) + w_2^2 \cdot \text{Var}(x_2) + 2w_1 w_2 \cdot \text{Cov}(x_1, x_2)$

となる. データより,

$$\begin{cases} \operatorname{Var}(x_1) &= 0.000151099 \\ \operatorname{Var}(x_2) &= 0.000656448 \\ \operatorname{Cov}(x_1, x_2) &= 0.000231408 \end{cases}$$

が得られるため 2 , w_1 , w_2 を求めるためには,

$$Q(w_1, w_2) = 0.000151099w_1^2 + 0.000656448w_2^2 + 2 \cdot 0.000231408 \cdot w_1w_2$$

を最大にする最適化問題を解けば良いことになる. そこで, ラグランジュの未定乗数法を用いて, ラグランジュ関数を次のように定義する.

$$F(w_1, w_2) = Q(w_1, w_2) - \lambda(w_1^2 + w_2^2 - 1)$$

= 0.000151099 $w_1^2 + 0.000656448w_2^2 + 2 \cdot 0.000231408 \cdot w_1w_2 - \lambda w_1^2 - \lambda w_2^2 + \lambda$

これを w_1, w_2 で偏微分して0とおくと

$$\frac{\partial}{\partial w_1} F(w_1, w_2) = 2(0.000151099w_1 + 0.000231408w_2 - \lambda w_1) = 0$$

$$\frac{\partial}{\partial w_2} F(w_1, w_2) = 2(0.000231408w_1 + 0.000656448w_2 - \lambda w_2) = 0$$

したがって,

$$\begin{cases} 0.000151099w_1 + 0.000231408w_2 &= \lambda w_1 \\ 0.000231408w_1 + 0.000656448w_2 &= \lambda w_2 \end{cases}$$

この式を行列で表現すると

$$\begin{pmatrix} 0.000151099 & 0.000231408 \\ 0.000231408 & 0.000656448 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \lambda \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

となる. これは分散共分散行列 V の固有値 λ , 固有ベクトル w に相当する.

²ここでは $\mathrm{Var}(x_1),\mathrm{Var}(x_2)$ について不偏分散を用いている.

定義 2.1 (分散共分散行列 (2×2行列の場合)).

$$\begin{pmatrix} x_1 \text{ の分散} & x_1 \text{ と } x_2 \text{ の共分散} \\ x_1 \text{ と } x_2 \text{ の共分散} & x_2 \text{ の分散} \end{pmatrix}$$

記号で表すと

$$\begin{pmatrix} \operatorname{Var}(x_1) & \operatorname{Cov}(x_1, x_2) \\ \operatorname{Cov}(x_1, x_2) & \operatorname{Var}(x_2) \end{pmatrix}$$

 $w_1^2 + w_2^2 = 1$ の条件の下、Mathematica を用いて固有値を求めると、固有値は

$$\lambda = 0.0000611453, 0.000746402$$

となる.

2.3.1 第1主成分

最大固有値 $\lambda = 0.000746402$ に対する固有ベクトルは

$$w = -0.932057, 0.362312$$

である. したがって, 第1主成分は

$$p_1 = -0.932057x_1 + 0.362312x_2$$

となる。第1主成分の寄与率は、

$$\frac{0.000746402}{0.0000611453 + 0.000746402} = 92.4\%$$

であり、ここから第1主成分がサンプルの様子を良く表現できているということが読み取れる.

2.3.2 第 2 主成分

2番目に大きい固有値 $\lambda = 0.0000611453$ に対する固有ベクトルは

$$w = 0.362312, 0.932057$$

となる. したがって, 第2主成分は

$$p_2 = 0.362312x_1 + 0.932057x_2$$

となる. 第2主成分の寄与率は、

$$\frac{0.0000611453}{0.0000611453 + 0.000746402} = 7.6\%$$

である. なお, 第 2 主成分の係数ベクトルは第 1 主成分の係数ベクトルと直交している. また, 求めた係数ベクトルより, 確かに $w_1^2 + w_2^2 = 1$ が成り立つ.

3 主成分得点の持つ意味

3.1 因子負荷量による判断

定義 3.1 (因子負荷量). 因子負荷量とは, 因子分析において, 得られた共通因子が分析に用いた変数(観測変数) に与える影響の強さを表す値で、観測変数と因子得点との相関係数に相当する.

-1以上1以下の値をとり、因子負荷量の絶対値が大きいほど、その共通因子と観測変数の間に(正または負の)強い相関があることを示し、観測変数をよく説明する因子であると言える. (統計 WEB より)

表 1: 因子負荷量

	出塁率	長打率
第1主成分	-0.593	0.110
第2主成分	0.805	0.993

主成分にどのような意味付けがなされるかは、因子負荷量の大きさや符号から判断できる。因子負荷量が正(負)ならば、その変数が増えると主成分の値は増える(減る)。本研究で用いた出塁率と長打率では、因子負荷量は次のようになる。ここから、第1主成分は各球団が出塁率と長打率のどちらを強みとしているかを表し3、第2主成分は出塁率と長打率を総合して評価したものを表すものと考えることができる。

3.2 指標と得点の相関

"基本方針"の通り、「OPS」「NOI」「第1主成分得点」「第2主成分得点」の4指標それぞれについてチームの1試合平均得点との相関係数を計算したところ、以下の表の様になった。この表から、OPS、NOI、第2主成分とチーム平均得点との強い相関がみられたが、第1主成分とチーム平均得点との間には相関が無かったことがわかる。

表 2: チームの 1 試合あたりの平均得点との相関係数

指標	相関係数 r	主成分	相関係数 r
OPS	0.945	p_1	-0.181
NOI	0.939	p_2	0.931

4 まとめ

得られた主成分のうち、OPS のように「出塁率と長打率を総合的に評価するもの」として用いることができるのは第 2 主成分の方となった。では、果たして第 2 主成分は指標として有用であると言えるだろうか。OPS と比較した際に、僅かながら相関係数で負けている上に、第 2 主成分には複雑な係数が付いているため**簡便性では遙かに劣っている**と言える。その点、OPS は出塁率、長打率ともに係数が 1 であり、かつチームの平均得点と極めて高い正の相関を示していたことから、万能ではないながらも OPS がいかに優れた指標であるかを改めて感じることができる。

今後の課題

当初はセ・パ両リーグの規定打席到達者⁴の成績を対象とすることを考えたが、その場合規定打席未到達者による影響を完全に排除してしまうことになり、かといって全選手を平等に扱おうとすると打席数の違いによる誤差が生じかねないため、'17 芝浦祭の QS の研究と同様にチームの成績を対象とした。前回は失点と自責点の違いの処理を課題に挙げ、今回は規定打席未到達者の成績の処理方法が課題となったが、「野球の研究をする際にはチーム成績と個人成績の扱いを考えなければならない」ということについては前回と今回で共通していた。今後この課題の解決を図るにあたっては、必要に応じて多変量解析以外の分野を取り入れることを検討する必要があるように思う。

 $^{^3}$ 出塁率が増加すると p_1 が減少するため, p_1 が大きい球団は出塁率を強みとし, p_1 が小さい球団は長打率を強みとすることが分かる.

⁴規定打席:所属球団の試合数×3.1 (小数点以下四捨五入)

参考文献

- [1] http://npb.jp/: NPB.jp 日本野球機構, 最終アクセス日: 2018.5.18
- [2] http://ranzankeikoku.blog.fc2.com/blog-entry-2290.html:セパ年度別 打低打高早見表 [1950-2017]-日本プロ野球 RCAA&PitchingRun まとめ blog, 最終アクセス日:2018.5.10
- [3] https://ja.wikipedia.org/wiki/: Wikipedia—野球の各種記録, 最終アクセス日:2018.5.15
- [4] https://bellcurve.jp/statistics/glossary/660.html: 統計 WEB—統計用語集—因子負荷量, 最終アクセス日: 2018.5.19
- [5] 森田 浩 著, 図解入門ビジネス 多変量解析の基本と実践がよ~くわかる本, 秀和システム, 2014
- [6] 室 淳子, 石村貞夫 著, Excel でやさしく学ぶ多変量解析 [第2版], 東京図書株式会社, 2007
- [7] データスタジアム株式会社 著, 野球 × 統計は最強のバッテリーである セイバーメトリクスとトラッキング の世界, 中央公論新社, 2015

5 補足資料

次頁の表に以下の情報を掲載した.

- 1. NPB の公式 HP に掲載されている'06~'17 のペナントレースにおける各球団の打撃成績から引用したチームごとの出塁率と長打率、及びチームの得点の1試合平均
- 2. 1から計算した OPS と NOI
- 3.1から本研究で導出した第2主成分得点 (p_2 とする)

表 3: チームの出塁率と長打率 ('12~'17)

球団	出塁率	長打率	得点	OPS	NOI	p_2	球団	出塁率	長打率	得点	OPS	NOI	p_2
	'17 パ・リーグ						717 セ・リーグ						
ソ	.331	.421	4.46	.752	471	0.512	広	.345	.424	5.15	.769	486	0.520
西	.332	.420	4.83	.752	472	0.512	神	.327	.371	4.12	.698	451	0.464
楽	.324	.390	4.09	.714	454	0.481	De	.311	.391	4.17	.702	441	0.477
オ	.316	.380	3.77	.696	443	0.469	巨	.318	.373	3.75	.691	442	0.463
日	.313	.357	3.56	.670	432	0.446	中	.300	.365	3.41	.665	422	0.449
口	.297	.351	3.35	.648	414	0.435	ヤ	.306	.338	3.31	.644	419	0.426
'16 パ・リーグ						'16 セ・リーグ							
日	.340	.385	4.30	.725	468	0.482	広	.343	.421	4.75	.764	483	0.517
ソ	.341	.386	4.42	.727	470	0.483	巨	.310	.384	3.60	.694	438	0.470
ロ	.326	.363	4.05	.689	447	0.456	De	.309	.385	3.97	.694	437	0.471
西	.335	.395	4.30	.730	467	0.490	神	.312	.308	3.51	.620	415	0.400
楽	.324	.368	3.78	.692	447	0.460	ヤ	.331	.378	4.13	.709	457	0.472
オ	.317	.355	3.47	.672	435	0.446	中	.309	.353	3.47	.662	427	0.441
		'15 パ	・リー	グ					'15 セ	・リー	グ		
ソ	.340	.408	4.52	.748	476	0.503	ヤ	.322	.377	3.99	.699	448	0.468
日	.330	.378	4.27	.708	456	0.472	巨	.313	.354	3.40	.667	431	0.443
ロ	.320	.368	3.90	.688	443	0.459	神	.317	.343	3.23	.660	431	0.435
西	.335	.406	4.38	.741	470	0.500	広	.312	.368	3.51	.680	435	0.456
才	.321	.354	3.60	.675	439	0.446	中	.313	.344	3.28	.657	428	0.434
楽	.311	.338	3.22	.649	424	0.428	De	.306	.373	3.53	.679	430	0.459
		'14 バ	・リー	グ					'14 セ	・リー	グ		
ソ	.344	.396	4.22	.740	476	0.494	巨	.321	.391	4.14	.712	451	0.481
オ	.334	.382	4.06	.716	461	0.477	神	.335	.376	4.16	.711	460	0.472
日	.321	.379	4.12	.700	447	0.470	広	.337	.420	4.51	.757	477	0.514
口	.314	.378	3.86	.692	440	0.466	中	.325	.364	3.96	.689	446	0.457
西	.329	.384	3.99	.713	457	0.477	De	.317	.383	3.94	.700	445	0.472
楽	.327	.356	3.81	.683	446	0.450	ヤ	.339	.412	4.63	.751	476	0.507
		'13 バ	リー	グ					'13 セ	リー	グ		
楽	.338	.379	4.36	.717	464	0.476	巨	.326	.400	4.15	.726	459	0.491
西	.331	.363	3.96	.694	452	0.458	神	.326	.358	3.69	.684	445	0.452
口	.331	.374	3.97	.705	456	0.469	広	.319	.367	3.87	.686	441	0.458
ソ	.339	.409	4.58	.748	475	0.504	中	.315	.359	3.65	.674	435	0.449
オ	.323	.366	3.56	.689	445	0.458	De	.325	.390	4.38	.715	455	0.481
日	.326	.368	3.71	.694	449	0.461	ヤ	.327	.377	4.01	.704	453	0.470
'12 パ・リーグ						'12 セ・リーグ							
日	.315	.363	3.54	.678	436	0.452	巨	.326	.367	3.71	.693	448	0.460
西	.316	.353	3.58	.669	434	0.444	中	.311	.334	2.94	.645	422	0.424
ソ	.304	.349	3.14	.653	420	0.435	ヤ	.325	.361	3.47	.686	445	0.454
楽	.307	.332	3.41	.639	418	0.421	広	.297	.330	2.97	.627	407	0.415
口	.320	.350	3.47	.670	437	0.442	神	.302	.316	2.85	.618	407	0.404
オ	.301	.337	3.08	.638	413	0.423	De	.298	.320	2.93	.618	405	0.406