

# MLB 選手の年俵に影響を与えるセイバーメトリクスの有意性

2020sc057 新實健大

指導教員：河野浩之

## 1 はじめに

MLB では、セイバーメトリクスを用いて戦術が立てられる。セイバーメトリクスとは、野球をデータから客観的に、統計学的に分析する手法である。その目的として、新たな視点から旧来の指標では明らかに出来なかった選手の本当の貢献度を計測することにある。旧来の分析指標には選手の能力を正しく評価できない問題があった。MLB に対して、NPB では近年データ分析の重要性が認識されつつあるが、いまだに旧来の成績指標を用いており、MLB と比較するとかなりの格差があるという問題がある。

また、NPB ではどのようなプレーが年俵の査定に反映されているか説明がないことが原因で年俵に不満を持つ選手の数は多い [1]。そこで本研究は、NPB 球団に向けセイバーメトリクスの有意性を示し、選手の年俵決定の新たな視点を提案し、普及を促す。

## 2 多変量解析による年俵推定の先行研究

多変量解析による年俵推定の先行研究を表 1 に示す。本間ら [2] は、読売ジャイアンツの各選手の成績と年俵額のデータから、年俵額を決定する要因の分析と、具体的な評価を数式として示した。

本間らの研究では、(次期年俵額 = 年俵係数 × 前年度の年俵) と仮定した。渡部ら [3] は、関東地区のセ・リーグの 3 球団の比較を行い、読売ジャイアンツの選手の年俵に影響を与える要因分析を行った。課題として有意係数の数値が高い点、選手データの年数の幅が狭いことを挙げた。

表 1 年俵の要因分析の先行研究

	分析手法	結果
本間ら [1]	重回帰分析	勝率の重みが最大。 投球回数の重みが最小。
渡部ら [2]	重回帰分析	外国人、試合数、FA 権 の要因が影響ある。 有意係数の数値が高い。

## 3 年俵予測モデルの提案手法

本研究の分析手順を図 1 に示す。本研究では、各野手の成績指標を説明変数とし、年俵を目的変数とする重回帰分析を行う。使用データには、セイバーメトリクス指標を用いたデータと旧来の成績指標のみのデータの二つのデータを用いて分析を行う。

また前シーズン 50 試合以上出たロサンゼルス・エンゼルスの一部の選手を対象とし、現在までの MLB におけるすべてのキャリア成績を分析対象とする。また渡辺ら [3]

の課題である、多重共線性に注意し相関係数が 0.95 以上の相関の強い変数と分散拡大係数 VIF が 10 以上の変数を調整することで有意確率が高くなるように分析した。その後、分析結果を基に二つのモデルの精度の比較を行う。選手のデータサイトと使用したツールを表 2 に示す。

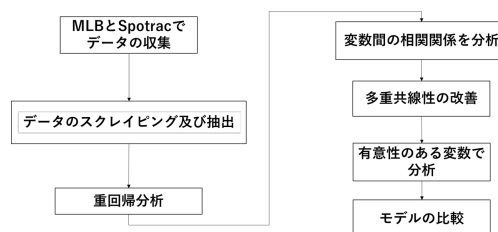


図 1 本研究の分析手順

表 2 本研究に用いたデータサイト・ツール

サイト・ツール	概要
MLB	公式データサイト 詳細なデータが揃う
spotrac	年俵データの抽出
Excel	データ管理と整理
Exploratory	分析ツール
Adobe Acrobat	OCR ソフトウェアテキスト認識

## 4 変数の選定について

相関関係を調べ、まとめられた変数で VIF の計算を行った。相関係数が 0.95 以上の相関の強い変数と分散拡大係数 VIF が 10 以上の変数を選定する。また 3B (三塁打数)、CS (盗塁死)、IBB (敬遠) は 1 シーズンでの数が少なく、年俵への影響が小さいと分析し、それら変数を削除した。

### 4.1 旧来の成績指標の変数の選定

相関係数が 0.95 以上であった AB, TB, RBI, R, H の変数を G にまとめ、どの説明変数も VIF が 10 以下になるように、AVG (打率)、2B (二塁打数)、SO (三振) の削除を行った。重回帰分析で用いた説明変数は、G (試合数)、HR (本塁打数)、SB (盗塁)、OBP (出塁率)、SF (犠飛)、SLG (長打率)、BB (四球) である。

### 4.2 セイバーメトリクス指標の相関関係

相関係数 0.95 以上であった AB, TB, RBI, R, H の変数を G に、SLG を OPS にまとめた。またどの説明変数も VIF が 10 以下になるように、AVG (打率)、SO (三振)、BB (四球)、2B (二塁打数) の削除を行った。重回帰分析で用いた説明変数は、G (試合数)、HR (本塁打数)、SB

(盗塁), OBP (出塁率), SF (犠飛) とセイバーメトリクス指標の OPS, GO/AO, P/PA, BABIP である。

## 5 分析結果及びモデルの比較

### 5.1 モデルの性能を示す指標

本研究では、モデルの性能を示す指標として、決定係数 ( $R^2$ ) を用いる。決定係数  $R^2$  は、重回帰分析におけるモデルの適合度を表す指標。この値は 0 から 1 の範囲で、1 に近いほどモデルが観測データによく適合していることを表す。決定係数  $R^2$  の計算式は以下の通りである。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

### 5.2 旧来の成績指標と重回帰分析の結果と考察

変数の選定前のモデルと選定後のモデルでは、決定係数  $R^2$  が  $R^2 = 0.2824$  から  $R^2 = 0.419649$  に向上した。これは選定前のモデルでは説明できなかった約 13 % の観測データを新たなモデルでは説明できることを示す。これよりモデルがより高精度になったことが分かる。また選定後の重回帰分析で  $p$  値が 0.05 以下になった有意な変数は BB (四球), G (試合数), SB (盗塁) である。中でも G (試合数) の  $p$  値は 0.0187 であり、試合に多く出ることが年俸に大きく影響することが分かった。しかし、決定係数が  $R^2 = 0.419649$  と 41 % の実数値しかモデルで説明できず、タイトルや長期契約, FA 権の要因が考えられる。

### 5.3 セイバーメトリクス指標と重回帰分析の結果と考察

変数の選定前のモデルと選定後のモデルでは、決定係数  $R^2$  が  $R^2 = 0.310899$  から  $R^2 = 0.643732$  に向上した。これは選定前のモデルでは説明できなかった約 30 % の観測データを新たなモデルでは説明できることを示す。重回帰分析で一部の係数の係数及び  $t$  値,  $p$  値を表 3 に示す。 $p$  値が 0.05 以下になった有意な変数は HR (本塁打数) のみであった。分析結果から HR を打つ選手をほど年俸が高くなりやすいと言える。 $p$  値が 0.05 以下の説明係数が少ない原因として、長期契約の選手もデータに含まれるからと考えられる。またすべてのセイバーメトリクス指標がモデルへの有意性を示せるわけではないことも分析された。

表 3 セイバーメトリクス指標と回帰係数

変数	回帰係数	t 値	p 値
BABIP	-21727899.53	-0.768	0.446
GO/AO	-3310706.963	-1.47	0.148
HR	496030.1511	2.15	0.0373
OBP	88920499.44	1.89	0.065
OPS	4386104.314	0.285	0.777
P/PA	9261625.67	1.72	0.0919

### 5.4 セイバーメトリクス指標と旧来の成績指標のモデルの比較

二つのモデルの決定係数の比較を表 4 に示す。二つのモデルの決定係数を比べるとセイバーメトリクス指標を含めたモデルのほうが約 22 % の観測データを多く説明できるモデルと分かった。これより、セイバーメトリクス指標の有意性があると考えられる。

表 4 作成した二つのモデルの比較

データ	決定係数 $R^2$
旧来の成績指標	0.419649
セイバーメトリクス指標を含む	0.643732

## 6 結び

本研究では、エンゼルス野手のキャリア成績を基に、重回帰分析を行い、年俸を予測するモデルを旧来の成績指標とセイバーメトリクス指標を用いた二つのモデルを作成し、モデルの比較を行った。旧来の成績指標として G (試合数) を含む 17 個の成績指標を用い、セイバーメトリクス指標には OPS, GO/AO, P/PA, BABIP の 4 つを用いた。それぞれ二つのモデルで変数の選定を行い、重回帰分析を行った結果、決定係数が旧来の成績指標では約 0.42、セイバーメトリクス指標のが約 0.64 と精度が高いことが分かった。これにより、セイバーメトリクス指標の有意性を分析できた。

しかし、セイバーメトリクス指標を用いた重回帰分析結果、 $p$  値が 0.05 以下を示す指標が HR (本塁打数) のみで、どのセイバーメトリクス指標が年俸に有意な影響を与える可能性大きいかまで断言することができなかったが、長期契約の選手を考慮しなかった点や本研究では複雑化するため考慮しなかったタイトル獲得など成績以外の要因も含めて分析することで精度の向上が考えられる。

## 7 参考文献

- [1] 日本プロ野球選手会, <https://jpbpa.net/research/>, (2023-1-24).
- [2] 本間 宏利, 石井 裕子, “多変量解析によるプロ野球選手の次期年俸推定 (投手編)”, 釧路工業高等専門学校紀要, NO.39, pp.33-38, (2005).
- [3] 渡辺静香, 朝日弓未, “日本プロ野球選手の年俸に影響を与える要因分析”, 日本計算機統計学会シンポジウム論文集, pp1-2, (2016).
- [4] ESSENCE OF BSEBALL, [https://1point02.jp/op/gnav/glossary/gls\\_index.aspx](https://1point02.jp/op/gnav/glossary/gls_index.aspx), (2023-12-12).
- [5] MLB, <https://www.mlb.com/>, (2023-12-12).
- [6] spotrac, <https://www.spotrac.com/>, (2024-1-12).