

日本人メジャーリーガーの統計的分析

2005MM030 北島康博

指導教員：田中豊

1 はじめに

メジャーリーグでは、セイバーメトリクスという統計手法を数十年前から用いて球団の構成、選手の評価を行っている。日本プロ野球では、この統計手法に対して認識は大きくなく、「打率」、「本塁打」、「打点」などの従来の指標を参考に評価しているのが実情です。本研究では、メジャーリーグで活躍する日本人選手に注目し、従来の指標と、セイバーメトリクスを用いて分析を行った。今シーズン日本選手の中で規定打席に到達した選手は、タイパベイズに所属する岩村明憲選手と、シアトル・マリナーズに所属するイチロー選手のみであった。また、今シーズンの成績と比べるためにイチロー選手のみ2004年の成績についても用いることとする。

2 データについて

データについては、「メジャーリーグベースボールMLB.com」[3]から2008年度のメジャーリーグ打率トップ100の選手の「試合数」、「打数」、「得点」、「安打」、「2塁打」、「3塁打」、「本塁打」、「打点」、「塁打数」、「四球」、「三振」、「盗塁」、「出塁率」、「長打率」、「打率」の15項目を用いた。また、セイバーメトリクスにおいては、「OPS」、「RC27」、「IsoD」、「PP/BB」、「P/PA」、「PA/K」、「IsoP」、「BIBAP」の8項目を用いた。セイバーメトリクスのデータについては[2]を参考に計算を行った。

3 従来の指標の主成分分析

主成分分析とは多変量のデータの合成変量を考えて、少ない次元でデータを把握する統計的方法である。2008年度のメジャーリーグデータを主成分分析することにより、どの選手がどんな能力があって、どんなタイプの選手なのかを把握する。

3.1 主成分分析

主成分は、累積寄与率が80パーセント以上になるところまで用いることにする。第一主成分寄与率(36.8%)、第二主成分寄与率(19.9%)、第三主成分(14.7%)、第四主成分(寄与率9.9%)となった。

3.1.1 主成分の解釈

第1主成分では、ほぼ全てが負の方向に働いており、これは総合的な力を意味している。そして、負の方向にいけばよく活躍し、正の方向にいけばあまり活躍してないといえる。よって、第1主成分は「年間を通して活躍しているかどうか」を表す軸である。第2主成分では、「本塁打」、「長打率」、「出塁率」が大きく正の方向に働いており、「打数」、「安打」、「試合数」、「三塁打」が大きく負の方向に働いている。つまり、「パワーヒッタータイプか先頭バッタータイプか」を表す軸である。第3主成分では、「打率」、

「盗塁数」、「出塁率」、「四球」が大きく正の方向に働いており、「打点」、「本塁打」、「三振」が大きく負の方向に働いている。よって「思いっきりがよく積極的に打っていく打者か、選球眼良くチャンスを作る打者か」を表す軸である。第4主成分では、「盗塁」、「三塁打」、「三振数」が大きく正の方向に働いており、「打率」、「安打」が大きく負の方向に働いている。よって「足は早いが三振数の多い打者か、打って出塁するタイプか」を表す軸である。

3.1.2 散布図

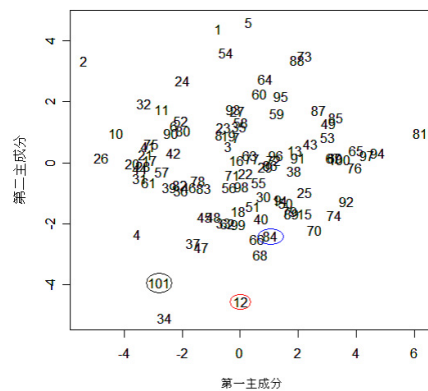


図1 第1主成分得点と第2主成分得点の散布図

12がイチロー選手(2008年),84が岩村選手,101がイチロー選手(2004年)

3.1.3 散布図の考察

イチロー選手の第一主成分は、ほぼ中央の位置になっている。これは、平均的な成績を残したことを示している。第二主成分は、大きく負になっている。これは、メジャーリーグの中でもずば抜けて先頭バッターとしての特徴がでている。今季のイチロー選手は平均的な成績であったが、先頭バッターとしての特徴が非常に強くでていることがわかる。また、2004年度と比較して、全体の成績はかなり落ちているが、イチロー選手の打撃特性はほとんど変わっていないことがわかる。岩村選手は第一主成分ではやや正の位置になっている。これは、平均より悪い成績を残したことを示している。第二主成分では、やや負の位置になっている。これは、先頭バッターとしての特徴がイチロー選手ほどではないが、十分でている。

4 従来の指標のクラスター分析

クラスター分析とは、対象間の距離を定義して、距離の近さによって対象を分類する方法のことである。樹形図(テンドログラム)を使いそれぞれの打者の特徴を分類し

ていく。第一主成分得点から第四主成分得点を用いた。そして、本研究では、解析の際にワード法を用いた。

4.1 樹形図

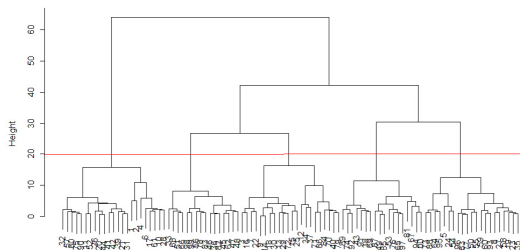


図 2 2008 年デンドログラム
イチロー選手は 12, 岩村選手は 84

4.2 分析結果

- 樹形図の左から、
- 第 1 群:非常にすばらしい成績を残している。メジャーの中でもオールスター級の選手の集まり。
 - 第 2 群:1 年間ほとんど休まず試合にでたタフな選手の集まり。
 - 第 3 群:長打力はないが、三振数が非常に少なく、足の速い打者の集まり。
 - 第 4 群:試合数が少なく、全体の中でも一番成績の低い選手の集まり。
 - 第 5 群:成績は標準的。しかし試合数が他のグループと比べ低いことから、1 年間を通して活躍できていない選手の集まり。

5 セイバーメトリクス

セイバーメトリクスとは従来の指標では野球選手の成績は選手の能力だけでなく試合状況や運に左右されてきた。それでは、純粋に選手の能力だけを評価したらどうなるのか。それを実現してくれるのがセイバーメトリクスです。セイバーメトリクスの目的は、旧来からある運と能力がないままになった指標の不公平を解消すること。そして、選手の真の実力を評価することです。

5.1 指標の説明

- OPS:得点との関連性が高い出塁率と長打率を単純に足したもの。セイバーメトリクスにおいて打者を評価する数値として用いられることの多い指標。
- RC:選手の活躍を得点に貢献した値として評価する指標。
- RC27:その選手だけで打線を組んだ場合、1 試合どれだけの得点を生み出せるか。
- IsoD:安打以外にどれだけ出塁できたかを示す指標。
- PA/BB:フォアボール 1 個を選ぶのに必要な打席数が数値化できる指標。

P/PA:1 打席あたりに相手投手に何球投げさせるかという指標。

PA/K:三振をひとつ奪われるまでに何打席を要するかという指標。IsoP:純粋に長打の割合を測ることができる。長打率との違いは、単打をカウントされないところ。

BABIP:運にどれくらい左右されたのかの目安となる指標。これは、本塁打以外でフィールド内に飛んだ打球が、どれくらいの割合で安打になるかという指標である。

5.2 セイバーメトリクスのデータ

12(2008 年のイチロー選手),84(岩村選手),101(2004 年のイチロー選手), 平均は打率トップ 100 の平均

	12	84	101	平均
OPS	0.747	0.729	0.869	0.827
RC27	5.67	5.03	7.68	5.95
IsoD	0.051	0.075	0.042	0.067
PA/BB	14.69	10.1	15.55	13.03
P/PA	3.57	4.14	3.51	3.78
PA/K	11.52	5.4	12.1	7.36
IsoP	0.076	0.106	0.083	0.177
BIBAP	0.337	0.339	0.401	0.322

5.3 考察

イチロー選手:OPS は 7 割り前後で平均的,また得点貢献度は低い値となっている。2004 年と比べると、2 点ほどの下がっている。次に三振する確率は非常に低いが、一打席での投球数や安打以外での出塁率の低さから、三振数が少ないのはイチロー選手が非常に早打ちであると思われる。よって、選球眼があまり良いとは言えない。最後に BABIP が 2004 年と比べてかなり下がっていることから、今シーズンは運もなかったのかもしれない。

岩村選手:OPS は 7 割前後で平均的,また得点貢献度は低い値となっている。次に三振する確率は非常に高い,また一打席での投球数が非常に多く安打以外での出塁率が平均よりも高い。よってかなりの遅打ちであるが、追い込まれると三振する可能性がかなり高いことがわかる。

6 おわりに

本研究を行ってきて、セイバーメトリクスを使うことによって従来の指標ではわからない選手たちの特徴がわかった。また今シーズンは規定打席に達した日本人選手が二人しかいなかったことが非常に残念だった。来年は松井選手や福留選手など、他の日本選手についても分析してみたい。

7 参考文献

- [1] 伊藤竜・伊藤貴章氏:メジャーリーグの多変量解析 - イチロー選手の打撃分析, 南山大学数理情報学部数理科学科卒業論文,2006
- [2] 野球の見方が 180 度変わるセイバーメトリクス, サンケイ総合印刷株式会社,2008
- [3] メジャーリーグベースボール MLB.com <http://mlb.mlb.com/index.jsp>