

メジャーリーグの多変量解析

— イチロー選手の打撃分析 —

2003MM031 伊藤竜 2003MM032 伊藤貴章

指導教員: 木村美善

1 はじめに

現在、メジャーリーグでは数々の日本人選手が活躍している。その中でも、イチロー選手はルーキーながら首位打者、最多安打、盗塁王という3つの主要タイトルを獲得してマリナーズの地区優勝に貢献し、新人王とアメリカンリーグMVPにも選ばれた。記録を残し続けるイチロー選手をメジャーリーガーの打率トップ30の選手と比較することによって、どのようなタイプの選手であるかを分析するために本研究に取り組んだ。

なお、伊藤竜は主に主成分分析と因子分析を、伊藤貴章は主に回帰分析とクラスター分析と調子の波を担当した。

2 データについて

主成分分析、クラスター分析、重回帰分析に用いたデータについては、「メジャーリーグベースボールMLB.com」[6]から2004年度のメジャーリーグ打率トップ30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。

因子分析に用いたデータは、2004年度のアメリカンリーグとナショナルリーグそれぞれの打率トップ30の選手の「打数」、「得点」、「安打」、「2塁打」、「3塁打」、「本塁打」、「打点」、「四球」、「三振」、「盗塁」の10項目を用いた。

3 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35.6%)、第二主成分(寄与率24.2%)、第三主成分(寄与率16.2%)、第四主成分(寄与率8.3%)となった。

3.1 主成分の解釈

第一主成分では、「本塁打」、「打点」、「長打率」が大きく正の方向に働いており、「打数」、「安打」、「盗塁」が負に大きく働いている。つまり、「パワーヒッターか先頭バッター」を表す軸である。第二主成分では、「三振」を除く全ての項目が負の方向に働いている。その中でも、「試合数」、「打数」、「安打」が特に大きく働いている。そこから、「荒っぽい打者が、確実性のある打者か」を表す軸である。第三主成分では、「2塁打」、「打点」が大きく正の方向に働いており、「四球」、「出塁率」、「打率」が大きく負の方向に働いている。したがって、「思い切りがよく積極的に打っていく打者か、チャンスを作る打者か」を表す軸である。第四主成分では、「2塁打」、「打率」が大きく正の方向に働いており、「3塁打」、「三振」、「盗塁」が大きく負の方向に働いている。以上より、「打って出塁するタイプか、足は速いが三振の多い打者か」を表す軸である。

3.2 散布図

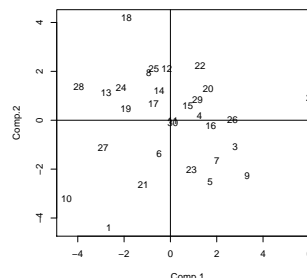


図 1: 第一主成分と第二主成分の散布図

3.2.1 散布図の考察

打率の良い選手の方から順に番号を1から30としており、1がイチロー選手である。

3.2.2 図1について

横軸が第一主成分となっており、イチロー選手は負の位置になっている。これは、本塁打や長打率があまりないことを示している。縦軸は第二主成分で、負にいくほど安打が多く、正にいくほど三振が多いという特徴を表している。イチロー選手は第二主成分も負になっている。負ということは、ヒットの確率が高いことを表している。この2004年のイチロー選手は首位打者に輝き、安打が他の選手と比べ、群を抜いて多いことがわかる。このデータは、セクション2で説明したように打率が高い選手を集めているので、打率が高い選手はパワーヒッターが多いこともわかる。図1を見ると、イチロー選手の特徴が顕著に表れている。第一主成分と第二主成分を合わせて見ても、イチロー選手は他の選手とは外れている。これは、メジャーリーガーにも少ない大打者であるといえる。

3.2.3 第三主成分と第四主成分について

第三主成分と第四主成分についても同様に散布図をつくった。イチロー選手は横軸が第三主成分で、負になっている。これは、よくチャンスを作ることを意味する。縦軸は第四主成分で、2塁打、打率が高いほど正になり、イチロー選手は正になっていることから、三振が少なく打って出塁するバッターであることがわかる。第三主成分と第四主成分を合わせて見ると、やはりイチロー選手はよく出塁してチャンスを作る先頭バッター向きであるということになる。

3.2.4 2つの散布図から

イチロー選手は打率が高いメジャーリーグのトッププレーヤーの中でも、際立っていることがわかる。

4 因子分析

4.1 アメリカンリーグ

第5因子の固有値が1より大きく、第4因子までの累積寄与率が80%にならないので第5因子までを用いた。

因子	第1因子	第2因子	第3因子	第4因子	第5因子
打数	0.089	0.875	0.202	-0.115	0.177
得点	0.346	0.488	0.152	-0.028	0.783
安打	0.018	0.989	0.063	-0.082	0.071
2塁打	0.401	0.052	-0.350	0.431	0.344
3塁打	-0.167	0.162	0.930	-0.024	-0.063
本塁打	0.759	0.073	-0.308	0.312	0.361
打点	0.915	0.155	-0.180	0.189	0.244
四球	0.349	0.011	-0.156	0.297	0.388
三振	0.245	-0.227	-0.025	0.939	0.025
盗塁	-0.521	0.258	0.648	-0.176	0.153
固有値	2.184	2.160	1.629	1.342	1.136
寄与率	0.218	0.216	0.163	0.134	0.114
累積	0.218	0.434	0.597	0.731	0.845

表 1: アメリカンリーグ因子負荷量

表1は因子数5でバリマックス回転後の結果を表したものである。第1因子の中で負荷量が大いのは、「2塁打」、「本塁打」、「打点」である。したがって、この因子は「長打力のある強打者」であると考えられる。このように、第2因子は「ミート力があり、確実性の高い打者」、第3因子は「走力」、第4因子は「荒っぽさ」、第5因子は「オールラウンドな打者」と解釈できる。

4.2 因子得点の散布図

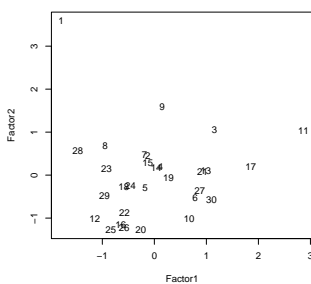


図 2: アメリカンリーグ因子得点の散布図

図2は横軸を第1因子得点、縦軸を第2因子得点として30選手をプロットしたものである。1がイチロー選手である。

4.3 考察

図2の横軸は第1因子である。横軸を見ると、イチロー選手は大きくマイナスになっている。第1因子で正にいる選手は、打点と本塁打が大きく関わっている。しかし、負にいるからといってパワーが無いわけではない。盗塁が多い選手は負になるので、イチロー選手は負に大きく働いたと考えられる。次に縦軸の第2因子を見てみる。イチロー選手は他の選手から大きく離れている。第2因子はミート力が優れた選手ほど正の値が大きくなるのである。横軸を第

3因子得点、縦軸を第4因子得点とした散布図ではイチロー選手は横軸がやや負の位置になっており、これは2塁打の影響だと考えられる。縦軸の第4因子もイチロー選手は負になっている。したがって、イチロー選手は先頭バッターに向いていると考えられる。イチロー選手は、アメリカンリーグの中でもずば抜けたミート力があるという結果になった。三振が多く大振りなバッターが多いアメリカンリーグの中では、目立つ存在ではないだろうが。

4.4 ナショナルリーグ

イチロー選手はアメリカンリーグに所属しているので、ナショナルリーグの打率上位29選手にイチロー選手を加えた30選手で分析を行った。アメリカンリーグと同様にそれぞれの因子の解釈をすると、第1因子は「オールラウンドな打者」、第2因子は「ミート力があり、確実性の高い打者」、第3因子は「走力」、第4因子は「荒っぽさ」と解釈される。

4.5 因子得点の散布図

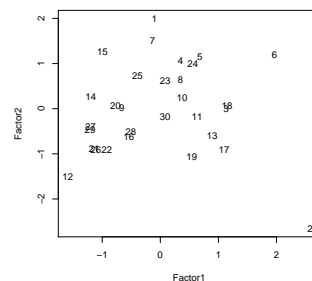


図 3: ナショナルリーグ因子得点の散布図

図3は横軸を第1因子得点、縦軸を第2因子得点として30選手をプロットしたものである。1がイチロー選手である。

4.6 考察

まず初めに、図3を見てみる。横軸は第1因子となっており、正にいくほど強打者である。しかし、イチロー選手はやや負になっている。これは打数が大きく影響しており、先頭バッターであるので、負になったと考えられる。縦軸である第2因子は、正に働くほどミート力があることになる。イチロー選手は1番大きく正の位置にあり、ナショナルリーグでもミート力は1番であり、十分に通用する選手であると考えられる。次に、第3因子得点と第4因子得点の散布図を見る。横軸である第3因子では、正に働くほど走力に優れているということである。イチロー選手は正の値が大きいの走力に優れている。これは盗塁の多さが影響していると考えられる。縦軸である第4因子は正に働くほど大振りを意味する。イチロー選手は三振が少なく、やはり安打が多いという結果になった。

4.7 全体のまとめ

図2と図3を比較する。2つの散布図の縦軸はミートを意味する。アメリカンリーグでは、イチロー選手は1人だけずば抜けて大きな値をとっている。しかしナショナルリーグでは、イチロー選手に近い位置に他の選手もいる。このことから、アメリカンリーグではナショナルリーグよりもバ

ワータタイプの選手が多い。そして、ナショナルリーグはアメリカンリーグよりもミート力がある選手が多いということが言える。図3よりも図2の方がイチロー選手が際立っている。したがって、イチロー選手はアメリカンリーグの方が活躍できると考えられる。

5 調子の波の激しさとは

5.1 調子の波とは

実際はコンスタントに成績を残す打者でもチャンスとか運によって波が激しく振舞うように見えることがあり得る。したがって、実際にその選手が、調子の波の激しい打者なのか、調子の波のない（コンスタントに成績を残す）打者なのかは分からない。この分析では、選手の真の調子の波について分析していく。この分析は参考文献[1]を参照し行っていく。データについては「スポーツナビ」[5]から用いた。

分析に使用するデータ

- 移動平均 - 移動平均グラフは、短い期間における打率を効果的に表現する方法である。この分析では、8ゲームごとの移動平均とグラフを使用する。
- MAX - MIN - 最も高い移動平均と最も低い移動平均の差を求めることにより、どれくらいの波の激しさがあるのかを計算する。この統計量をMAX - MINと呼ぶことにする。
- 良いゲームと悪いゲームの連 - ある選手の打率が0.279であったとする。そこで、4打数2安打であった場合を調子の良いゲームとして+で示す。また、もし4打数1安打であった場合は調子の悪いゲームだとして0で示す。調子の良い日や悪い日が連続する興味深いパターンを連と言ひ、それは、悪いゲームが連続する調子の波(0000のような)、もしくは良いゲームが連続する調子の波(++++のような)と同じである、また7ゲーム以上連続している場合を長いストリークと定義する。
- 良く打てる日とあまり打てない日 - ノーヒットのゲーム数と4本以上ヒットを打ったゲーム数を数える

コンスタントに成績を残す打者(Constant)

毎打席のヒットの確率を $P = 0.335$ (イチロー選手のメジャーリーグにおける平均的な打率に近似)と仮定する。そこで、2004年シーズンのイチロー選手の全ゲームの結果をこの打率を用いて50人シミュレーションする。

波の激しい打者(Streaky)

好調時の打率を $P_H = 0.435$ 、不調時の打率を $P_C = 0.235$ とする。また、波の激しい打者は、あるゲームにおいて好調であれば、次のゲームにおいても好調を維持する確率が高い。波が激しいためには、この確率が0.9のように大きな値をとらなければならない。これは同様の状態が続く確率が高いということを意味する。Constantな打者と同じように50人シミュレーションする。

表2を見てみると、イチロー選手の成績はConstantモデルとよりも、Streakyモデルの方が類似性が高いようだ。言い換えると、このStreakyモデルはConstantモデルよりもイチロー選手のデータによく一致している。

表 2: イチロー選手とConstant・Streakyの平均値

	MAX - MIN	ストリーク	連	0本	4本
イチロー-03	0.447917	2	72	35	5
イチロー-04	0.428571	1	72	27	10
イチロー-05	0.336812	2	74	36	2
平均(Constant)	0.381157	1.44	81.36	31.9	3.8
平均(Streaky)	0.472003	2.44	73.82	34.16	7.1

5.2 判別分析

判別分析は、個体(あるいは対象)がどのグループに属するかが明確であるデータを用いて判別モデルを構築し、そのモデルを用いて所属不明の個体がどのグループに帰属するかを判別する方法である。([4]参照)

マハラノビスの距離

この方法では、標本が各グループの平均点までどれだけ距離があるかを調べ、すでに判明している最も短い距離のグループに属するように判定する。

表 3: マハラノビスの距離

	Constant	Streaky
イチロー-03	3.74167	1.098452
イチロー-04	24.38046	5.851923
イチロー-05	4.23585	5.958058

線形判別分析

Constant・Streakyの各50人ずつのデータから判別関数を求める。判別関数は次のようになる。

$$f = 7.06898799x_1 - 0.01943518x_2 - 0.07112459x_3 + 0.02639219x_4 + 0.26160600x_5 + 0.2432864$$

判別関数式の中の x_1, x_2, x_3, x_4, x_5 はそれぞれMAX-MIN、長いストリーク、連、ノーヒット、ヒット4本以上を示す。

誤判別率は0.14(14%)であった。 f 値と各グループへの所属のp-値は次のようになった。

表 4: 所属のp-値

	fの値	Constant	Streaky
イチロー-03	0.4815197	$5.87e^{-01}$	0.965952
イチロー-04	1.4610961	$1.83e^{-04}$	0.376102
イチロー-05	-1.204552	$5.16e^{-01}$	0.377726

5.3 考察

データの統計値と判別分析より2004年シーズンのイチロー選手は調子の波の激しい打者に分類される可能性が高い。メジャーリーグ最多安打を記録した2004年シーズンのイチロー選手は2003年や2005年と比べるとシーズン中の成績の残し方が少し違うということが分かる。特に1日に4本以上ヒットを打つゲーム数とノーヒットのゲーム数に特徴が出ている。通常Streakyな選手は4本以上ヒットを打つゲーム数とノーヒットのゲーム数がそれぞれ多くなる。2004年のイチロー選手はStreakyな選手同様、4本以上ヒットを打つゲーム数が多い。しかしノーヒットのゲーム数は少ないという結果が出ている。これが、最多安打を残せた要因の一つであると考えられる。イチロー選手は自分

の打撃の調子をよく理解し、不調なときには不調なりの打撃、好調なときにはヒットを量産することを心がけて打席に立っていたのだろう。その結果、ノーヒットの試合は少なくなり、ヒット4本以上の試合が多くなったと考える。自分の好不調の波をよく理解しているようだ。しかし、調子の波のより精度の高い結論を導くためには、より多くのシミュレーション（各1000人程度）が必要だと考える。

6 打率の予測と要因

この分析において、打率を予測する回帰式を導く。回帰式により予測される打率はどのような説明変数から得られるか、また2004年のイチロー選手の打率0.372とどのくらいの差が生じるのかを打率を目的変数として分析していく。なお、イチロー選手の打率をほかの選手の成績から得るために、打率トップ30の選手からイチロー選手自身を除いた29選手のデータを使用して分析していくこととする。また、「塁打数」は多重共線性の問題から、あらかじめ除くこととする。この分析では、あらかじめ「試合数」、「打数」、「安打」、「出塁率」を除く9項目を用いた。なぜなら、安打数と打数以外にどの説明変数が打率の予測に大きな影響をおよぼすのかを調べるためである。また、変数選択にはRのstep関数を使った。([2]参照)

表 5: 出力結果

	回帰係数	標準誤差	t-統計量	p-値
定数項	0.2472000	0.02058000	12.014	$2.16e^{-11}$
3塁打	-0.0010460	0.00060550	-1.728	0.097347
本塁打	-0.0007733	0.00035540	-2.176	0.040074
三振	-0.0001899	0.00007843	-2.422	0.023740
盗塁	0.0003373	0.00023570	1.431	0.165784
長打率	0.2015000	0.04799000	4.200	0.000342

標準誤差は0.009164, 決定係数は0.6432, F検定（自由度5, 23）の結果は8.293, 有意水準は0.0001344となり、表5より回帰式は次のようになる。

$$y = 0.2472 - 0.001046x_1 - 0.0007733x_2 - 0.0001899x_3 + 0.0003373x_4 + 0.2015x_5$$

x_1 は3塁打, x_2 は本塁打, x_3 は三振, x_4 は盗塁, x_5 は長打率を表している。

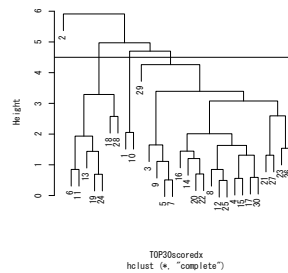
6.1 考察

表5に出てくる5種類の変数が打率の予想に大きく影響しているものである。打率の高い選手の中にはイチロー選手のような安打を積み重ね出塁する先頭打者タイプの選手と高い長打率を持ち本塁打を量産するパワーヒッタータイプの選手が存在する。割合で考えるとパワーヒッターのほうが多い傾向にあり「3塁打」「本塁打」「三振」「長打率」が選ばれたのだと考えた。また、3塁打, 本塁打, 三振はバッターが大振りであると起こりやすい結果なので打率予測では負の要因となっている。メジャーリーグで打率の高い選手は強打者が多いということが分かった。1番打者などは打席数が多いので打率は上がりにくいようだ。

7 クラスタ分析

樹形図（デンドログラム）を使いそれぞれの打者の特徴を分類していく。主成分分析で求めた主成分得点を第四主成分まで使用する。（[3]参照）

7.1 樹形図(デンドログラム)



分析結果

第1群 長打率がずば抜けて高い。他に類をみないパワーヒッターである。

第2群 標準的な選手であると考ええる。

第3群 先頭打者タイプの選手であると考ええる。

第4群 多く長打を期待でき打点が高い。よってクリーンナップを任される選手が多いと考える。

考察

イチロー選手と同群に分類された選手は1人だけだった。今回使用した打率トップ30の選手はクリーンナップなどを任される強打者が多いことが分かる。打席の多く回ってくる1・2番打者で高い打率を残せる選手は少ないようだ。イチロー選手のようなタイプの打者が常に3割以上の高打率を残すことはあまりないことだ。

8 おわりに

メジャーリーグ挑戦以来、常に3割以上の成績を残しているイチロー選手の安打を量産できる要因やメジャーリーグにおける打者タイプも明らかになった。メジャーリーグにはイチロー選手のようなタイプで好成績を残せる打者は少ないという結果がでた。

謝辞

本研究を進めるにあたり、御指導頂いた木村美善教授及びゼミ関係者の友人に深く感謝致します。

参考文献

- [1] J. アルバート・J. ベネット 後藤寿彦 監修 加藤貴昭 訳: メジャーリーグの数理科学 上, シュプリンガー・フェアラーク東京, 2005.
- [2] 間瀬茂・神保雅一・鎌倉稔成・金藤浩司: 工学のためのデータサイエンス入門, 数理工学社, 2004.
- [3] 岡田聡・山口正氏: 多変量解析法による日米野球選手の分析, 南山大学数理情報学部数理科学卒業論文要旨集, 2003.
- [4] 田中豊・脇本和昌: 多変量統計解析法, 現代数学社, 1983.
- [5] スポーツナビ,
<http://sportsnavi.yahoo.co.jp/index.html>
- [6] メジャーリーグベースボール MLB.com,
<http://mlb.mlb.com/NASApp/mlb/index.jsp>