

プロ野球投手の統計的分析

2001MM101 山本将大

指導教員 木村 美善

1 はじめに

私は先発投手の成績に注目し、現在中日ドラゴンズのエースである川上憲伸に興味を持った。そこで、過去3年間の川上憲伸の成績を集めて、球数、打者数、被安打数、被本塁打数、失点、四死球数、三振数、内野ゴロ数などを、途中で降板した試合も一人で9回まで投げたものに変換して解析した。また、2005年度先発投手として10試合以上登板している投手を調べ、投球内容が、川上憲伸と似た投手を探してみた。

2 データについて

重回帰分析に用いたデータについては、過去3年間分の日中スポーツプロ野球結果 [5] より、 Y =失点、 X_1 =球数、 X_2 =打者数、 X_3 =被安打数、 X_4 =被本塁打数、 X_5 =四死球数、 X_6 =得点を許したイニング数、 X_7 =奪三振数、 X_8 =内野ゴロの数、 X_9 =被塁打数、 X_{10} =味方の得点、 X_{11} =相手の順位の11項目の記録を用いた。そして、相手順位以外の10項目の記録は、1人で9回まで投げたものに変換した。なお、目的変数 Y は防御率とした。表1はこれらの一部を表したものである。クラスター分析に用いたデータについては、昨年度の日中スポーツプロ野球結果 [5] より、先発投手として10試合以上登板している69名分の打者数、被安打数、三振数、四死球数、失点をそれぞれ球数で割ったものと、球数を用いた。

表1 重回帰分析のデータの一部

失点	味方 得点	被本塁 打数	三振数	被塁 打数	四死 球数	打者数
0	4	0	7	6	0	31
10.8	3.6	3.6	7.2	25.2	1.8	45
2.1	5.2	0	4.2	7.3	3.1	34.3
0	11.6	0	5.2	7.7	1.3	34.3
7.7	0	1.3	9	19.4	1.3	40
2.5	2.5	0	7.4	12.3	3.9	36.9
1	7	1	10	4	0	28
5.7	9	1.1	4.5	15.8	2.5	36.2
2.6	3.9	0	3.9	14.2	0	37.4
3.4	2.3	1.1	5.7	15.8	0	35
4.5	1.1	1.1	5.7	19.2	0	35
2.5	2.3	1.1	6.8	12.4	0	39.6
5.2	3.9	2.6	10.3	21.9	2.5	40
1.2	2.3	0	8.2	7	1.3	37.4

3 解析方法

3.1 線形回帰モデル

ある目的変数 Y とそれに影響すると考えられる P 個の説明変数 X_1, X_2, \dots, X_p の間の関係式を求め、それにもとづいて X_1, X_2, \dots, X_p の値から Y の値を予測したり、その際の各 X_i の寄与の大きさを評価したりする分析である。説明変数が1つの場合を単回帰分析、2つ以上の場合を重回帰分析という。線形回帰モデルは、

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi} + u_i \quad i = 0, 1, \dots, n$$

$\beta_0, \beta_1, \dots, \beta_p$ は回帰係数、 u_i は誤差項。 ([1], [2] 参照)

3.2 クラスタ分析

クラスター分析とは、異なるものの混ざりあっている対象を、それらの何らかの意味で定義された類似度を手がかりにして似たものを集め、いくつかの均質的なものの集落に分類する方法である。 ([4] 参照)

4 重回帰分析の結果と考察

変数選択法より回帰分析を行った結果、平成17年度、平成16年度、過去3年分のデータについて、それぞれ次の結果が得られた。

表2 平成17年度回帰分析結果

定数項	係数	標準誤差	T 値	P 値
X_1	-0.00176	0.0356	0.0493	0.9614
X_4	-0.04428	0.6589	0.0672	0.9474
X_8	-0.05105	0.3339	0.1529	0.8808
X_{10}	-0.1236	0.1507	0.8205	0.4267

表3 平成16年度回帰分析結果

定数項	係数	標準誤差	T 値	P 値
X_1	0.10904	0.04367	2.4970	0.0247
X_4	0.05669	0.8377	0.0677	0.9469
X_6	0.1777	0.52413	0.3391	0.7392
X_{10}	0.15863	0.10406	1.5244	0.1482

表4 過去3年間分の回帰分析結果

定数項	係数	標準誤差	T 値	P 値
X_1	0.03368	0.0266	1.2646	0.2121
X_3	0.00223	0.00331	0.6757	0.5025
X_4	-0.59400	0.46301	1.2829	0.2057
X_6	-0.09907	0.29032	0.3413	0.7344

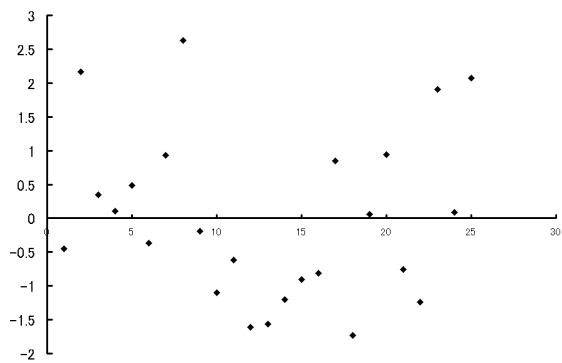


図1 平成17年度の失点に関する標準化残差プロット

平成17年度(表2)の結果においては球数、被本塁打、内野ゴロ数、味方の得点が説明変数として選択された。これらの係数はいずれも負であり、防御率が良くなる方向に影響するが、その影響は味方の得点を除くと小さい。三振数は変数として選択されず、失点にあまり影響していないと言える。これからわかる川上憲伸の特徴は三振が多く奪えるほど調子が良い日は球に勢いがあるために、コースが多少甘くても打ち取れると感じ、投球が雑になる。したがって、甘いコースにきた球を相手バッターに本塁打を打たれる。その後も球威があるので本塁打を打たれたことを気にもとめずに投げ続けるケースが多いと考えられる。逆に調子が悪いと感じているときは、球に勢いはないもののコースを厳しく狙って投げるので、球数が多くなるが、相手バッターの凡打が多くなり、失点が少なくなると考えられる。また、味方が点を取るほど川上憲伸は失点しにくい結果になった。図1は \hat{Y} に対する標準化残差プロットであるが、特定のパターンが見られるものの、外れ値は見られない。

平成16年度の結果(表3)については、すべての回帰係数が正であり、球数と味方の得点が防御率悪化への影響が大きくなっている。この結果から川上憲伸の調子が良いときほど、相手打者は早いカウントから打ちに行き、凡打を打つケースが多いと考えられる。被本塁打数は、失点にはあまり影響していない。これは川上憲伸が速球派の投手であり、調子の善し悪しに関係なく、ストレートに的を絞られ、本塁打を打たれるタイプの投手であるとわかる。このように平成16年度は平成17年度とかなり異なることが分析結果からもわかり、興味深い。

3年間分を分析した結果(表4)では、被本塁打数が防御率が良くなる方に最も強く影響する結果となった。また、被安打数、球数は防御率が悪くなる方へ強く影響している。逆に被本塁打数は平成17年度、平成16年度と同様にあまり反映しない結果となった。この結果から川上憲伸は、16年度の考察で述べたように、被本塁打数は調子の善し悪しにあまり関係ないが、球数が多い日は被安打数も多くなり、大量失点するので早いイニングで降板させたほうが良いと言える。この意味でも川上憲伸にとって平成16年度は他の2年とはかなり違ったと考えられる。(〔3〕参照)

5 クラスタ分析の結果と考察

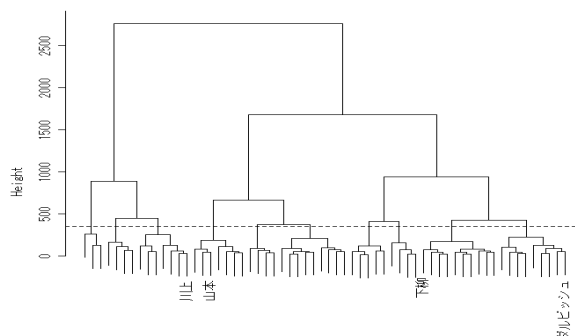


図2 投手69人のデンドログラム(最長距離法)

この結果(図2参照)から阪神タイガースの福原、横浜ベイスターズの門倉、広島東洋カープの黒田、大竹が川上憲伸とかなり近い特徴を持つと考えられる。この4名の特徴は、各チームの投手陣の柱として1年間ローテーションを守り、三振も比較的多く奪うことができ、四死球数が少ないことである。さらに決め球をストレートとすることが多い、よく言われる強気なピッチングをする傾向がある。また、中日ドラゴンズの山本、阪神タイガースの下柳などのベテラン投手や、中日ドラゴンズの中田、北海道日本ハムファイターズのダルビッシュなどといった新人投手とはかなり異なった結果となった。

6 おわりに

本研究では被本塁打数、三振数は失点にあまり反映されておらず、逆に球数、味方の得点は大きく影響している。これからわかる特徴としては、三振が多く奪えるほど調子が良い日は球に勢いがあるために、コースが多少甘くても打ち取れると感じ、投球が雑になるという結果となった。さらに、平成16年度が川上憲伸にとって特別な年であったことも確認できた。また、クラスタ分析の結果として川上憲伸と特徴が似ていた阪神タイガースの福原、横浜ベイスターズの門倉、広島東洋カープの黒田、大竹の成績にも注目して、来シーズン以降川上憲伸の成績に注目していきたい。

参考文献

- [1] 伊藤哲二・入倉則夫・奥野忠一・片山善三郎・上郡長昭・藤原信夫：工業における多変量データの解析，日科技連出版社（2002）
- [2] 永田靖・棟近雅彦：多変量解析法入門，(2002)
- [3] 西脇崇：500mlペットボトルの売り上げに関する統計的分析，南山大学数理情報学部数理科学科卒業論文（2004）
- [4] 佐和隆光：回帰分析，朝倉書店（2002）
- [5] 中日スポーツ：(株)中日新聞社