

プロ野球の勝敗に関する統計的分析

2015SS021 伊藤 諒

指導教員：白石高章

1 はじめに

私は小、中学生の頃に野球、大学ではソフトボールをしていた。スポーツでは、前評判で弱いと言われているチームが、強豪チームに勝つことがある。強豪チームはどのような戦力、戦略で勝ち上がっているのか。また、弱いチームには、強豪に勝つために強豪にない戦略があると考えられるので、統計的分析を行い少しでも明らかにしていく。また、野球には”セオリー”というものが存在しているが、果たしてそれが勝利に近づくうえで”正しい”セオリー”であるのかをデータを用いて分析し、勝利に対してどのような要素が大きな影響を与えているかを統計的に分析をする。

2 データ

本研究では、プロ野球セ・リーグ(以後、セ)とパ・リーグ(以後、パ)の2013年から2017年までを対象とした。1年毎の試合データ5年間を12球団分[4],[5]用意して行う。参考にする試合は、交流戦を含めたシーズンを通しての試合のみである。オープン戦、クライマックスシリーズ、日本シリーズのデータは使用しない。

本論で活用するデータは、各球団の

- x_1 :1シーズンの勝率
- x_2 :1試合あたりの平均1塁打数
- x_3 :1試合あたりの平均2塁打数
- x_4 :1試合あたりの平均3塁打数
- x_5 :1試合あたりの平均本塁打数
- x_6 :1試合あたりの平均盗塁数
- x_7 :1試合あたりの平均犠打数
- x_8 :1試合あたりの平均四球数
- x_9 :1試合あたりの平均死球数
- x_{10} :1試合あたりの平均得点
- x_{11} :1試合あたりの平均失点
- x_{12} :1シーズンのピタゴリアン期待値
- x_{13} :1試合あたりの平均被安打数
- x_{14} :1試合あたりの平均被本塁打数
- x_{15} :1試合あたりの平均与四球数
- x_{16} :1試合あたりの平均与死球数
- x_{17} :1試合あたりの平均奪三振数
- x_{18} :1試合あたりの平均暴投数
- x_{19} :1試合あたりの平均ボーク数
- x_{20} :1試合あたりの平均失策数
- x_{21} :1シーズンの守備率
- x_{22} :1シーズンの総得点
- x_{23} :1シーズンの総失点

である。本論では、勝利には攻撃時と守備時のどのようなデータが大きく影響しているか、などといったことを分析

していく。

3 分析方法

分析方法として、ピタゴリアン期待値、重回帰分析、クラスター分析を行った。重回帰分析では攻撃の変数と守備の変数で目的変数を変え、さらに変数減少法を用いた。クラスター分析は、全球団の1シーズンをまとめて分析して行った。

4 ピタゴリアン期待値

1チームの年間の得点と失点を用いて計算する(鳥越規夫[3])、式は以下である。

$$\text{ピタゴリアン期待値} = \frac{(\text{得点})^{1.64}}{(\text{得点})^{1.64} + (\text{失点})^{1.64}}$$

ピタゴリアン期待値は近年発見されたセイバーメトリクス的一种である。単純な式ではあるが、実際のデータを代入すると、当てはまりがよいということになり、使用されている。全12球団の5年間の勝率とピタゴリアン期待値の相関係数は0.931である。

表1 ピタゴリアン期待値

年度	セ平均(勝)	セ平均(ビ)	パ平均(勝)	パ平均(ビ)
2013年	0.489	0.491	0.512	0.511
2014年	0.507	0.494	0.501	0.509
2015年	0.490	0.492	0.510	0.501
2016年	0.492	0.491	0.508	0.510
2017年	0.497	0.494	0.503	0.502

表1はピタゴリアン期待値の平均を年度ごとに、セとパに分けて計算したものである。差は0.01以内でほとんど収まっている。平均得点と平均失点を用いたこの式が勝率と高い相関関係がある事もわかる。また、セとパでは平均得点と平均失点において大きな差があり、パとセの総合的な戦力においても大きな差があると考えられる。

5 重回帰分析

表2は、 x_{10} を目的変数とし、 x_{10} に関係するであろう攻撃時の $x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9$ の8のデータを説明変数とし、重回帰分析を行った(菅民郎[1])。変数減少法により、 x_{18}, x_{19}, x_{20} が減少した。決定係数は0.893、自由度修正済み決定係数は0.880である。四球が平均得点に関して、説明変数の中でも高い相関関係がでており、p値も限りなく0に近い値となっている。四球は一塁打と同じで、塁を一つ進めるといふことには変わりはないが、投手のリズムが狂うことで、得点へ繋がっていると考えられる。”セオリー”と呼ばれる犠打には重回帰分析ではほとんど相関係

数がないという結果がでた。

表2 x_{10} を目的変数とした重回帰分析

	回帰係数	標準誤差	p 値
x_2	0.491	0.056	5.96×10^{-12}
x_3	0.243	0.147	0.103
x_4	1.4.3	0.422	0.002
x_5	1.154	0.132	7.09×10^{-12}
x_8	0.402	0.064	6.87×10^{-8}
x_9	0.570	0.261	0.033

表3 x_{11} を目的変数とした重回帰分析

	回帰係数	標準誤差	p 値
x_{13}	0.553	0.047	2.0×10^{-16}
x_{14}	0.984	0.128	3.35×10^{-10}
x_{15}	0.514	0.064	8.77×10^{-11}
x_{16}	0.864	0.237	0.6×10^{-3}
x_{17}	0.064	0.033	0.058
x_{21}	-18.898	7.387	0.013

表3は、 x_{11} を目的変数とし、 x_{11} に関係するであろう守備時の $x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}, x_{21}$ の9のデータを説明変数とし、重回帰分析を行った(菅民郎 [1])。変数減少法により、 x_{18}, x_{19}, x_{20} が減少した。決定係数は0.925、自由度修正済み決定係数は0.917である。被安打、被本塁打が高い相関関係があるのは想定内である。やはり、与四死球が失点に与える影響は大きい。失策などの仲間のミスよりも投手自身のミスの方が精神状態に影響を及ぼし、失点に繋がっていると考えられる。

6 クラスター分析

1年ごとに12球団で $x_1, x_{12}, x_{22}, x_{23}$ 以外を使用してクラスター分析を行う(金明哲 [2])。全体を考察してみた結果、群分けされたものと成績を照らし合わせてみると、平均得点が高い群。平均得点と平均失点が似た値の群。平均得点が低く、平均失点が高い群となった。平均得点の高い群は四球、二塁打、犠打が多くなっている。相手の動揺する心理にうまく漬け込んで得点を重ねている。

平均得点と平均失点が似た値の群は、塁打の数が平均より多くても、被安打が多い。平均失点が高い群は、塁打が少なく、被安打が多くなっている。Deの平均得点と平均失点に着目してみる。平均失点が高いが、平均得点が毎年のように上がるとともに、良い群へ変化している。

7 考察

平均得点の高さが、平均失点の低さよりも、勝利への影響度が高いという結果と言える。また、平均失点は守備のエラーはあまり関係がなく、ほとんどが投手自身の影響が大きい。野球は投手8割、のセオリーは正しかった。

野球の”セオリー”と言われる作戦の中で、一番定番である犠打に得点への相関がほとんどないという結果がでた。近年フライボール革命により、犠打をする球団が減少してきた。実際に犠打を少なくする作戦自体は正しかったと言える。しかし、犠打を試合後半に使うことで、中継ぎ投手の1点も取られてはいけない。という心理にうまく漬け込み、精神状態を揺さぶるという点においては有効である。試合序盤から何度も犠打をするのではなく、相手にプレッシャーがかかる場面での犠打のサインに変わってきている。また、打ち勝つ野球より、守り勝つ野球がセオリーと言われている。今回の結果だけでは判断はできない。近年のフライボール革命もそうだが、打高投低の時代である。守り勝つ野球は絶対的な”セオリー”であるとは言えず、打ち勝つ野球が”セオリー”に変化しているのかもしれない。ピタゴリアン期待値が新しい指標として出るくらい、

勝利には、やはり得点と失点が大きな影響を及ぼしている。打つという動作以外にも、四球など相手のミスの後に攻撃側は集中して攻撃を畳掛ける。守備側は、四球や失策など集中力がかけてしまった後に立て直すことが重要である。

8 終わりに

昔は野球といえば、投手の投げ合いでロースコアの試合が多かったが、今では打ち合いがほとんどである。私たちのようにプロ野球観戦が好きな人にとっては、より楽しく野球を見ることが出来る。しかし、プレーをする側にとっては、投手は打たれることが多く精神的にきつい。打者は打って当たり前のように双方にとってもストレスの溜まる野球になってしまっているのではないかと思う。

参考文献

- [1] 菅民郎:『例題とExcel演習で学ぶ多変量解析 - 重回帰分析・判別分析・コンジョイント分析編-』オーム社出版、東京、2016年11月30日。
- [2] 金明哲『Rによるデータサイエンス』森北出版、東京、2007年10月25日。
- [3] 鳥越規央:『勝てる野球の統計学』岩波書店、東京、2014年3月12日。
- [4] 『NPB.jp 日本野球機構』
<http://npb.jp/>
1996年。
- [5] 『プロ野球データ Freak』
<https://baseball-data.com/>
2009年。