

プロ野球の打線と勝率の関連性に対する統計的分析

2018SS023 伊藤孝哉

指導教員：松田眞一

1 はじめに

2021年シーズンのプロ野球はヤクルトとオリックスが優勝した。この2チームは打線が強力であった。その理由は長打率と出塁率が高く、そしてそれらを有効的に生かす得点能力が高いためだと考えた。そこで有効的に生かされたかどうかを分割表で判断し、そこから重回帰分析を用いて、その結果からどの部門がどう効果的だったのかを分析する。そして優勝できなかったチームが次シーズン優勝するために必要な数字を予測する。

2 研究の流れ

まず分割表(金 [2] 参照)で分析する。そこで得られるP値と各球団の各月の出塁率と勝率をロジット変換する。勝率や出塁率といった比率にあたるすべての部門の数字をロジット変換するが長打率は打者の長打力を測る目安となる指標であって比率ではないのでロジット変換の対象にはならない。これ以降これら3つをロジット変換した値をそれぞれLP値、L出塁率、L勝率と表記する。LP値、L出塁率、長打率を説明変数に、L勝率を目的変数として重回帰分析(森棟 [3] 参照)を行う。今度は年間通してのLP値と長打率とL出塁率を説明変数に、1年間のL勝率を目的変数として重回帰分析を行う。そして、長打率と出塁率に関してどの部門を強化すべきなのかを再び重回帰分析で予測する。なお重回帰分析では減少法を用いた変数選択を行った。多重共線性が生じていないかはVIFで確認した。集まったモデル式から次シーズン優勝するために1球団例にして2021年シーズンの成績から優勝するための年間のP値の条件を予測する。

3 データについて

本研究ではスマートフォンアプリ『プロ野球速報 試合経過』[6]から分割表の度数を決める。そして各球団の月ごとのデータ(web[1] 参照)について、L勝率を目的変数、LP値、長打率、L出塁率を説明変数として重回帰分析を行う。また、年間のデータ(web[1] 参照)についても目的変数と説明変数が同様の重回帰分析を行う。集まったデータをもとに次シーズンの予測を行うが様々な要素の到達ラインの設定としてここ5年で1番強力打線と言われている2018年の西武が残した記録(web[5] 参照)を到達ラインとする。

4 分割表

分割表とは2つ以上の変数の間の関係を分析するための表である。

- ・長打も四球も出なかったインニング…(1)
- ・四球は出たが長打は出なかったインニング…(2)
- ・長打は出たが四球は出なかったインニング…(3)

・長打も四球も出たインニング…(4)

の4種類に分類して、さらに点数がとれたかとれなかったかの2つに分類して2×4の分割表を作成し、当てはまるインニングがあったらそれを度数として数えその累計を最終的に度数として月ごとに解析を行う。また年間通して月別で集めたものをすべて加えそれを度数として1年間どうだったかも解析する。

表1 3,4月の阪神タイガースのインニングでの得点の分割表

	(1)	(2)	(3)	(4)
得点有	22	16	27	11
得点無	48	25	22	3

表1は求めた分割表の一例である。この表に対してRでピアソンのカイ二乗検定を行うとP値が0.002696という値になった。このP値と有意水準 α を比較すると $P < \alpha$ となるので3,4月の阪神は四死球や長打を有効的に得点をとったと言える。そして、2021年シーズンで1番P値が低かったチーム(ヤクルト)と1番高かったチーム(DeNA)のP値と勝率、長打率、出塁率を表2にまとめた。

表2 ヤクルトとDeNAの勝率、P値、長打率、出塁率

球団名	勝率	P値	長打率	出塁率
ヤクルト	0.584	0.0247	0.397	0.333
DeNA	0.425	0.0868	0.402	0.320

ヤクルトは2021年優勝し、日本シリーズでも優勝したチームで逆にDeNAはリーグで最下位であり、全球団で一番勝率が低かった。長打率ではヤクルトよりも上回っているし、出塁率もヤクルトに劣ってはいるものの他の球団には引けを取らない。しかしP値の差が大きい。したがっていくら長打を打てたとしても、いくら塁に出たとしてもそれらを有効的に得点にしなければ意味がないという表れである。

5 解析結果とモデル式

本研究での解析結果をまとめる。まずは月別のL勝率とLP値、長打率、L出塁率の関係である。L勝率を目的変数に、LP値、長打率、L出塁率を説明変数にして重回帰分析を行いさらに変数選択を行った結果、LP値とL出塁率が残った、変数選択後の結果が表3である。

よってモデル式は以下の式(1)となった。L勝率を目的変数 y とし、LP値の説明変数を x_1 、L出塁率の説明変数を x_2 とする。

$$y = -0.2485x_1 + 1.0417x_2 - 0.0494 \quad (1)$$

表3 重回帰分析の結果1

	係数	標準誤差	t 値	P 値
定数項	-0.0494	0.4615	-0.107	0.915
LP 値	-0.2485	0.0382	-5.312	0.001
L 出塁率	1.0417	0.5022	2.074	0.042

そして今度は1年間の成績で先ほどと同様に行い変数選択をして結果すべての目的変数が残り結果は表4である。

表4 重回帰分析の結果2

	係数	標準誤差	t 値	P 値
定数項	-0.0790	0.7392	0.202	0.845
LP 値	-0.4430	0.0463	-9.574	0.001
長打率	1.7668	1.1546	1.963	0.055
L 出塁率	0.7999	0.5054	1.583	0.015

よってモデル式は以下の式(2)となった。L勝率を目的変数 y とし, LP 値の説明変数を x_1 , 長打率の説明変数を x_2 , L 出塁率の説明変数を x_3 とする。

$$y = -0.4430x_1 + 1.7668x_2 + 0.7999x_3 - 0.0790 \quad (2)$$

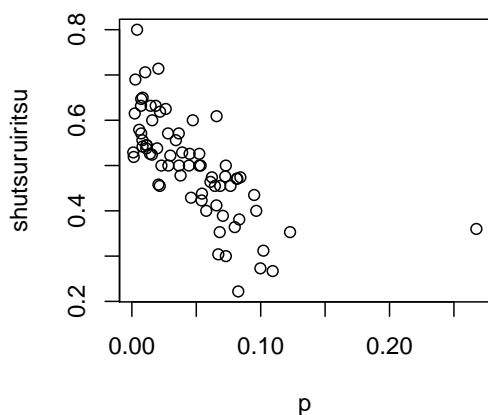


図1 全球団の月間のP値と勝率

月別で解析したら長打率の説明変数が消え, 年間の解析をしたら長打率の説明変数が残った。怪我や不振などで長打を打てる人がある期間いなかったとしてもチームとしては勝つことができるが, それでは1年間持たないので, 基本的には長打力が必要であることが言える。逆にある期間で長打力があっても必ず勝てる保証が無いという結果になったことが言える。出塁率の説明変数はどちらも残った。このことから長期的に見ても短期的に見ても得点をとるためには塁に出ることから始める必要があることが言える。月別のときの方が傾きが大きかったので, 特に優勝を争う大事な時期に出塁率は必須であると言える。そし

てP値についてだが, 月別のときも年間のときも説明変数は残った。やはり長打と四死球を有効的に生かす攻撃は常に必要であると言える。図1は全球団の月別のP値と勝率の散布図である。横軸がP値で縦軸が勝率である。この散布図からも有効的な攻撃ができているときほど勝率が高かったのは明らかであるのでそれを年間積み重ねた結果, 年間のP値が小さいチームが優勝するということと言える。年間積み重ねる中で, P値が大きく, 勝率が悪くなる時期が少しあったとしても短い期間であれば中解は可能であるので, 悪くなる期間をなるべく狭めるのがペナントレースでは重要である。

6 次シーズンへ向けて

解析で得たモデル式から優勝できなかったチームが優勝するために必要な数字をソフトバンクを例にして行う。1年間の最終の勝率と長打率と出塁率の目標ラインをそれぞれ0.580, 0.450, 0.350とした。年間のL勝率と長打率とL出塁率から1年間トータルで必要なP値をモデル式(2)から算出すると, 0.0346という値が予測された。P値が0.0346ということは2021年のヤクルト, 阪神, オリックス, ロッテよりも高い値である。このことから長打率と出塁率をここ5年間でトップクラスの数字まで磨き上げれば「長打と四死球」と「得点の有無」の2つの関係性がヤクルトなどよりも多少薄かったとしても優勝できると言える。

7 おわりに

結論としては優勝するためにはP値も長打率も出塁率も必要であることが分かった。今研究は攻撃のみを重点的に分析したが攻撃に囚われると守備力を失うリスクがあると解析中に思い, 選手の補強というのは我々のような素人が思っているよりも難しいものだった。選手補強が外れてしまうと前年よりもマイナス要素が大きくなるリスクもあるので特に攻撃と守備のバランスと選手を見極める力が最終的に大事になるだろう。

参考文献

- [1] データで楽しむプロ野球：
<https://baseballdata.jp/2/index.html> (2021年11月閲覧)
- [2] 金明哲：『Rによるデータサイエンス データ解析の基礎から最新手法まで』, 第2版。森北出版, 2017.
- [3] 森棟 公夫：『統計学 改訂版 (New Liberal Arts Selection)』, 有斐閣出版, 2015.
- [4] 日本野球機構：セ・パ公式戦 2022年試合日程
<https://npb.jp/games/2022/> (2021年12月閲覧)
- [5] 埼玉西武ライオンズ：2018年チーム成績
<https://www.seibulions.jp/game/stats/team/batter/2018.html> (2021年11月閲覧)
- [6] スマートフォンアプリ 『スポナビ 野球速報』 試合経過・順位表