

プロ野球の勝率とチーム成績の関連性に対する統計的分析

2020SS083 魚住 圭叶

指導教員：松田 真一

1 はじめに

私の趣味はプロ野球観戦である。どのチームもリーグ優勝を目指して143試合を戦っている。大学で統計を学び自分の好きなプロ野球でできることはないかと思い、プロ野球の勝率に関連しているチーム成績に興味をもち、研究に至った。また、2リーグ制であるためそれぞれのリーグ別に解析を行い、それぞれの違いも分析していきたい。

2 データについて

今回の研究ではプロ野球のデータのうち2016年から2023年までの8年分を使用した。

重回帰分析、クラスター分析ともにペナントレースにおける年度別のチーム成績(web[1]参照)を用いた。重回帰分析では、目的変数を勝率として説明変数に防御率、QS、BB/9、犠打、長打率、出塁率、チャンス時の安打、ピンチ時の被安打を用いて行う。クラスター分析では、重回帰分析をリーグ別に行い、セ・リーグ、パ・リーグで選択した変数、防御率、長打率、出塁率のペナントレースにおける年度別のチーム成績を用いて行う。

2.1 セパのルールの違い

日本のプロ野球には全部で12球団存在し、1リーグ6球団の2リーグ制で行われている。一方がセントラル・リーグ(以下セ・リーグ)、もう一方がパシフィック・リーグ(以下パ・リーグ)である。主なルール違いはパ・リーグでは、DH制を導入していることが挙げられる。DH制とは、投手が試合中に打席に立つことが無く、DHという守備に就かない打撃専門のポジションが追加される制度である。

3 解析方法

今回解析には、重回帰分析(丹慶[3]、涌井・涌井[4]参照)、クラスター分析(田中・脇本[2]参照)を用いた。クラスター分析ではウォード法を用いた。

4 重回帰分析の結果

プロ野球は、それぞれのリーグの最高勝率のチームがリーグ優勝となる。そこでセ・リーグ、パ・リーグ別に目的変数を勝率として、説明変数に成績を入れて重回帰分析を行い、AICを用いた変数減少法で変数選択して勝率に関連する成績を出した。

残差分析を行い、VIFを確認したが問題がなく正常な結果である。変数選択後の結果を表1、2に示す。

セ・リーグの決定係数は、0.8273、パ・リーグは、0.7998であった。係数、標準誤差は有効数字4桁、P値は小数点3桁で示す。P値0.001未満は0.000と示す。表1の得点圏安打はチャンス時の安打を意味する。

表1 変数選択後のセ・リーグの重回帰分析結果

	係数	標準誤差	P値
防御率	-0.1221	0.01322	0.000
QS	-0.0009234	0.0006012	0.132
長打率	1.181	0.2342	0.000
出塁率	1.007	0.6594	0.134
得点圏安打	0.0004837	0.0002271	0.0391

表2 変数選択後のパ・リーグの重回帰分析結果

	係数	標準誤差	P値
防御率	-0.09820	0.01107	0.000
長打率	1.406	0.2290	0.000
出塁率	1.769	0.4654	0.000

変数減少法での変数選択では、セ・リーグでは、防御率、QS、長打率、出塁率、チャンス時での安打、パ・リーグでは、防御率、長打率、出塁率が残った。セ・リーグとパ・リーグで違いが出た理由を考察する。

チャンス時での安打は、なぜセ・リーグでは選択されてパ・リーグでは選択されなかったのか。チャンス時での安打なので、チャンスに強いということではない、あくまでも得点圏に走者がいる状態での安打数なのである。これは、セ・リーグでは投手が打席に立つからだと考える。投手は基本的に打線の最後9番打者になることが多く、もちろん投手は他の野手と比べてバッティングは期待できない。つまり、バントを行える状況なら必ずバントを行う。また、状況によっては投手に代打を送る可能性もある。つまり、投手がバントをしてランナーを得点圏に送った状況下で安打を打って、投手から交代した代打の成績も良いチームの勝率が高くなっていると考えられる。パ・リーグでは、9番打者でも投手が打席に立つことなく、野手が打席に立つのでセ・リーグとは状況が違い、セ・リーグのみチャンス時の安打が選択されたと考えられる。

QSは、なぜセ・リーグでは選択されてパ・リーグでは選択されなかったのか。係数がマイナスになっているため、QS数が少ない方が勝率が高くなっている。リーグごとの投手の分業制への違いだと考えた。この結果は、セ・リーグではピッチャーの分業制が進んでいるからだと考える。リリーフに強力なピッチャーがいるほど先発投手からの交代が早く、リリーフ陣が強力な程、QSが少なくなると考える。また、投手が打席にたつので、打順の関係上パ・リーグよりもセ・リーグの投手交代が進んで行われやすいと考える。

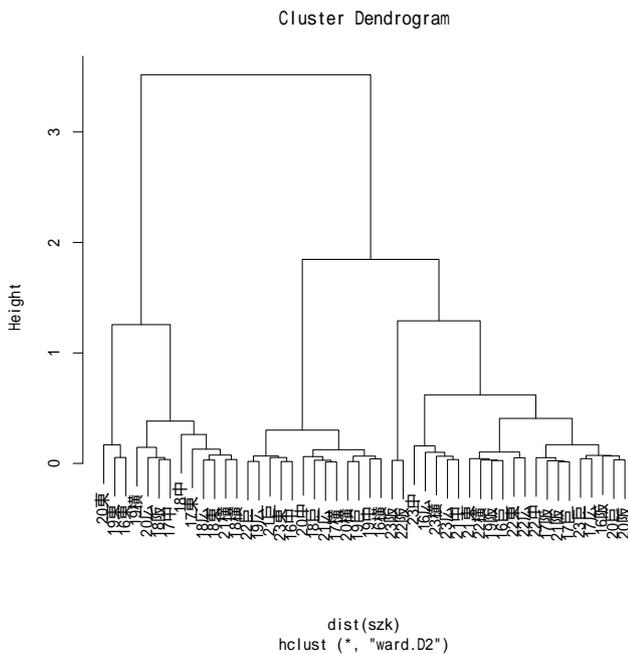


図1 セ・リーグのクラスター分析結果

5 クラスター分析の結果

重回帰分析を行い、変数減少法で選択した変数のうちセ・リーグ、パ・リーグで重複した変数(防御率, 長打率, 出塁率)を用いて, セ・リーグ, パ・リーグ別, セパ・リーグにクラスター分析を行った. 紙面の都合上, セ・リーグの結果のみ図1に示す.

5.1 セ・リーグの結果

図1を左から6群に分けた. 優勝チームが含まれる場合その年を横に記載する.

- 第1群 投手がとても劣っており出塁率が高いチーム
- 第2群 投手が劣っているが打撃が良いチーム 18
- 第3群 長打のみ優れているチーム 19
- 第4群 投手がとても優れていて長打がとても少ないチーム 23
- 第5群 投手が良くて打撃が最低なチーム 16
- 第6群 平均的なチーム 17, 20, 21, 22

5.2 パ・リーグの結果

- 第1群 平均的なチーム
- 第2群 投手がとても劣っていて, 打撃最高なチーム 18, 19
- 第3群 投手が劣っているチーム
- 第4群 投手が良く出塁率が低いチーム 17, 21
- 第5群 投手が最高で, 出塁率が低いチーム 20, 22, 23
- 第6群 投手がとても良く出塁率が高いチーム 16

5.3 考察

クラスター分析では, セ・リーグでは, 平均的なチームが優勝する年が多く, パ・リーグでは, 投手成績が良いチームが優勝することが多かった.

DH 制度の有無による差だと考える. DH 制度の無いセ・リーグの影響は3つあると考える.

1つ目はスタメン選手に打力が求められることだ. セ・リーグでは, パ・リーグよりも少ない8人で点をとる必要があるため, スタメンを決める際に打撃寄りの選手が重要視され, 投手戦よりも打ち合う試合になると考える.

2つ目は守備職人が起用されにくいことだ. セ・リーグでは守備力が優れた選手はスタメン野手の人数が少ない関係で起用されにくくなる. また, セ・リーグではDHのポジションが無いため, 外国人選手を守備に就かせる必要があり守備力が落ちることで点が入りやすくなると考える.

3つ目は投手への影響だ. セ・リーグでは試合中に投手が打席に立つ必要があり, インニング間の投球練習等がパ・リーグよりも調整が難しいと考える. また, 打席に立つことで打撃練習などの野手の練習にも参加する必要があり, 自身の投球練習のみを行うパ・リーグの投手と差ができる.

以上の理由から, セ・リーグでは投手で好成績を残しづらく, バランスの良いチームの勝率が高いと考える. パ・リーグではDH制度の影響で, 投手が好成績を残すことで, 投手成績が好成績のチームの勝率が高いと考える.

6 まとめ

重回帰分析での変数選択ではリーグごとに違いが出た理由は2つあると考える. 1つ目は, セ・リーグでは投手の分業制が進んでいるからだと考えた. 2つ目は, 投手が打席に立つかの違いによる差だと考えた. クラスター分析では, 優勝チームの傾向はDH制度の有無によってリーグごとに違いが出たと考えた.

7 おわりに

本研究を通して, 試合観戦だけではわからないことを, 解析を用いて今までとは違った視点からプロ野球を見ることができた. 主観的な分析ではなく, 数値を用いた客観的な分析ができ, 自身の趣味をより楽しめるようになった.

参考文献

- [1] えるてん:『nf3 プロ野球ヌルデータ置き場』.
<https://nf3.sakura.ne.jp/>
(2024年1月観覧).
- [2] 田中豊・脇本和昌:『多変量統計解析法』. 現代数学者, 1987.
- [3] 丹慶勝市:『図解雑学多変量解析』. ナツメ社, 2005.
- [4] 涌井良幸・涌井貞美:『図解でわかる多変量解析』. 日本実業出版社. 2001.