

# プロ野球 2 球団の統計的解析

2015SS040 森下凌

指導教員:白石高章

## 1 はじめに

私は幼い頃から野球観戦をすることが趣味で、特に攻撃面に関して、得点に結びつけるにはどのような要素が必要かといった点やチームの攻撃スタイルに関心を持っていた。複数のチームの比較をすることで、それぞれの「戦術」「得点に結びつけるための要因」といった特徴の違いを知ることができると思った。そこで、プロ野球界のセ・リーグとパ・リーグからそれぞれ、現在注目を浴び、最強球団との呼び声も高い 2 球団「広島東洋カープ」「福岡ソフトバンクホークス」(以後: 広島・ソフトバンク) を選んでデータ解析をしていくことにする。

## 2 データについて

本研究では、広島とソフトバンク 2 球団の 2005~2017 年までの 13 年間の公式戦のみのデータを使用する。データは、文献 [1], [2] による年間の得点、一塁打、二塁打、三塁打、本塁打、打点、盗塁、犠打、犠飛、四球、故意四球、併殺打の 11 要素と年度別順位を用いる。この 11 要素はすべて年間の合計を示しているため、各年の試合数で割り、1 試合あたりの平均値に変換してから分析に用いることとする。

## 3 分析方法

「得点」を目的変数とし、その他の 10 個のデータを説明変数として、広島とソフトバンク 2 球団それぞれの相関係数、重回帰を R によって分析し比較をする。2 球団それぞれの主成分散布図を R によって求め、比較をする。

## 4 相関係数

2 球団それぞれ別で相関係数行列を求め、得点に対する相関関係を一つの棒グラフにまとめた。

図 1 2 球団それぞれの得点に対する相関関係



一塁打、二塁打、本塁打、犠飛、四球、併殺打の 6 つは、2 球団に共通して正の相関が現れた。2 球団ともに  $-0.2$  以下の負の相関が現れた項目は故意四球あった。

## 5 重回帰分析

表 1 2 球団それぞれの重回帰分析結果

	ソフトバンク			広島		
	回帰係数	標準誤差	P 値	回帰係数	標準誤差	P 値
intercept	-0.14	4.11	0.76	-3.33	1.00	0.08
一塁打	0.35	0.48	0.54	0.77	0.16	0.04
二塁打	0.06	0.84	0.95	0.88	0.44	0.19
三塁打	-0.77	4.39	0.88	-1.1	0.76	0.29
本塁打	0.87	0.33	0.12	1.2	0.34	0.07
盗塁	-0.12	0.65	0.87	-0.23	0.49	0.67
犠打	1.28	1.02	0.33	0.65	0.39	0.24
犠飛	-2.59	6.01	0.71	4.04	0.91	0.05
四球	0.03	0.54	0.96	0.22	0.14	0.26
故意四	0.35	0.48	0.82	0.95	1.34	0.55
併殺打	2.2	1.13	0.19	-5.48	2.15	0.13

## 6 変数減少法

表 1 による重回帰分析結果の P 値に着目すると、広島では一塁打と犠飛以外の項目が 0.05 を上回っており、ソフトバンクではすべての項目が 0.05 を上回っている。このことから、文献 [3] に書かれている説明変数間の相関が強すぎて分析結果が不安定になる「多重共線性」とみなし、変数を選択して用いる必要があると考えた。その手段の一つである変数減少法から以下の分析結果を得た。

表 2 表 1 に変数減少法を用いた結果

	ソフトバンク			広島		
	回帰係数	標準誤差	P 値	回帰係数	標準誤差	P 値
intercept	-2.1	0.8	0.04	-2.89	0.56	0.01
一塁打	0.39	0.08	0.01	0.68	0.07	0.01
二塁打	—	—	—	0.77	0.3	0.05
三塁打	—	—	—	-0.79	0.5	0.19
本塁打	0.94	0.19	0.01	1.21	0.27	0.01
盗塁	—	—	—	—	—	—
犠打	1.41	0.25	0.01	0.55	0.24	0.87
犠飛	-3.38	1.31	0.04	3.61	0.54	0.01
四球	—	—	—	0.18	0.1	0.15
故意四	—	—	—	—	—	—
併殺打	2.51	0.51	0.01	-4.42	1.07	0.01

ソフトバンクに変数減少法を用いた結果、一塁打、本塁打、犠打、犠飛、併殺打を説明変数に置いたモデルが適切となった。標準誤差の大きいものを排除しながら回帰係数の数値が高い順に並べると併殺打、犠打、本塁打、一塁打となった。一方、広島に変数減少法を用いた結果、一塁打、二塁打、三塁打、本塁打、犠打、犠飛、四球、併殺打を説明変数に置いたモデルが適切となった。P 値を見て信頼性の薄い要素 (0.05 以上) を排除しながら回帰係数の数値が高い順に並べると犠飛、本塁打、二塁打、一塁打、併殺打となった。

## 7 主成分分析

文献 [4] を参考に R で分析を進め、累積寄与率が 8 割を越えたところで、関数 biplot による散布図を求め、各主成分を見出す。

図 2 広島 of biplot 散布図 (累積寄与率 0.83)

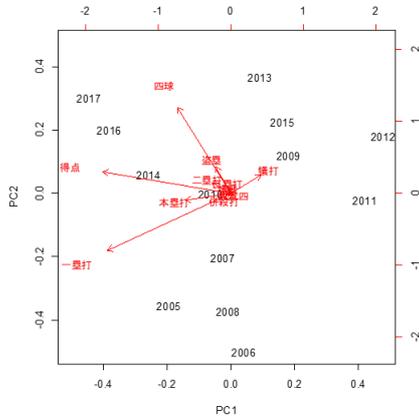


図 2 の PC1 に着目すると、右方向には犠打、故意四球が、左方向には得点、本塁打、一塁打が固まっている。従って、第一主成分は右へ行くほど「ヒットはなくとも泥臭い戦法でチームに貢献する (以後 A タイプ)」で、左へ行くほど「塁打で得点する (以後 B タイプ)」であると考えられる。PC2 に着目すると、下方向には微量ではあるが併殺打が、上方向には四球、盗塁が固まっている。従って、第二主成分は下へ行くほど「チャンスを生かせない (以後 C タイプ)」で、上へ行くほど「チャンスを生かす (以後 D タイプ)」であると考えられる。

図 3 ソフトバンク of biplot 散布図 (累積寄与率 0.95)

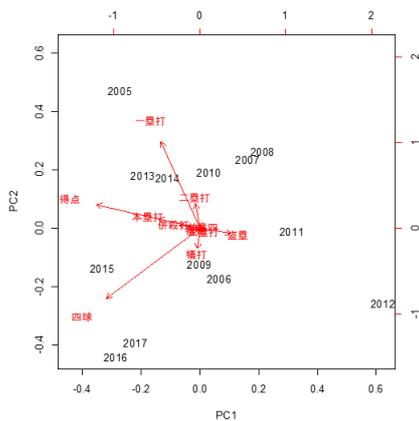


図 3 の PC1 に着目すると、右方向には盗塁が、左方向には本塁打、得点が固まっている。従って、第一主成分は右へ行くほど「小技でバッテリーを揺さぶる (以後 E タイプ)」で、左へ行くほど「B タイプ」であると考えられる。PC2 に着目すると、下方向には四球、犠打が、上方向には一塁打、

二塁打が固まっている。従って、第二主成分は下へ行くほど「A タイプ」で、上へ行くほど「ヒット量産 (以後 F タイプ)」であると考えられる。

## 8 広島 of 考察

重回帰で得た数値を見ていくと、犠飛の回帰係数は 3.61 と群を抜いて高い数値となっており、三塁にランナーを置いた場面でスクイズ等を行わず、振りぬくスタイルで得点していることがわかる。本塁打のほか、一塁打と二塁打も回帰係数は正で安定して得点に絡みついでおり、チャンスで一気に得点する強力打線ということが窺える。主成分分析を見ると、第一主成分の A タイプでは 2009, 2011, 2012, 2015 年, B タイプでは 2014, 2016 年, 第二主成分の C タイプでは 2005~2008 年, D タイプでは 2013, 2017 年があげられる。泥臭い戦法が目立つ年の順位は 5, 4 位と低く、本塁打が目立つ年は 3, 1 位と高い、チャンスを生かせない年は 6~4 位と低く、生かせる年は 3, 1 位と高い。以上を踏まえると、広島は平均して豪快な得点スタイルで、それに比例して順位も高くなり、泥臭い戦法が目立つときには順位が悪くなると考えられる。

## 9 ソフトバンク of 考察

重回帰で得た数値を見ていくと、本塁打で得点することを抜くと、犠打でランナーを進塁させる戦法が得点に影響している点が目立つ。主成分分析を見ると、第一主成分の E タイプでは 2011, 2012 年, B タイプでは 2013, 2015 年, 第二主成分の A タイプでは 2012, 2013 年, F タイプでは 2005, 2010, 2014 年があげられる。盗塁等の小技が目立つ年は 3, 1 位と高く、本塁打が目立つ年は 4, 1 位と落差がある。泥臭い戦法が目立つ年の順位は 1 位と高く、ヒットが目立つ年も 2, 1 位と高い。以上を踏まえるとソフトバンクは広島のように豪快に得点するケースは少なく、何かを犠牲にした泥臭い戦法が大きく得点に影響しているチームであり、この得点スタイルが目立つ年は順位も比例して高くなると考えられる。

## 10 終わりに

強いチームにも「強力打線」や「小技スタイル」と種類があり、強力打線のチームに小技スタイルが目立つと順位が低くなったりと、得点や順位変動の要因はチームによって様々であることが分かった。

## 参考文献

- [1] 『NPB.jp 日本野球機構』  
<http://npb.jp/>, 2018 年 6 月参照。
- [2] 『プロ野球データ Freak』  
<https://baseball-data.com/>, 2018 年 6 月参照。
- [3] 『多変量データ解析法 心理, 教育, 社会系のための入門』ナカニシヤ出版, 足立 浩平 [著], 2012 年。
- [4] 『R によるデータサイエンス』森北出版, 金 明哲 [著], 2007 年。