

プロ野球における得点力のあるチームの特徴

1.概要

どのような特徴のあるチームが得点をしやすいかを調べる。OPS、打率、本塁打数について、中央値以上の成績の選手の数と、第三四分位数以上の成績の選手の数のデータを使って**主成分分析**を行ったところ、**第一主成分は各項目の平均、第二主成分は本塁打数と打率のどちらの成績が良いか、第三主成分は他のチームと比べて成績が中央値以上の選手が多いか、第三四分位数以上の選手が多いか**を表していた。次に、これらの3成分を用いて**重回帰分析**を行った。そして、**打率型のチームよりも本塁打型のチームのほうが得点力が高く、打率型のチームの中では、抜群の成績の選手の数が多きチームの方が、本塁打型のチームの中では、より本塁打数の高いチームの方が得点力が高い**傾向にあるということが分かった。

2.背景・目的

プロ野球ではいろいろな特徴を持つチームがあり、また、年によってもチームの傾向が異なっている。今回は**攻撃面**に注目し、どのような特徴のあるチームが得点数が多いかを調べることにした。特に、**中央値以上の成績である選手の数と第三四分位数以上の成績である選手の数**という要素から研究した。

3.方法1(主成分分析)

用いたデータは**2017,2018,2019,2021年の日本プロ野球12球団**のデータである。2020年は新型コロナウイルス感染症の影響で試合数が減っていて、他の年と同様に扱えないため除いた。これらを合わせて48個のチームと考え、各チームの特徴を調べる。まず、各チームの中で**打席数の多い選手9人**を抽出する。この**9人のうち何人**が中央値、第三四分位数を超えているのかを、**本塁打数、打率、OPS^{*1}**についてそれぞれ調べる。この時の中央値や第三四分位数は、抽出した選手のなかで算出したものだ。これら調べた6つのデータに**標準化**(平均と分散が一定になるように値を調整する操作)をした後、**主成分分析^{*2}**を行い、チームの特徴はどのように端的にあらわされるのかを調べ、どんな特徴を持つチームの得点数が多い傾向にあるのかを調べる。

***1OPS**
出塁率と長打率を足し合わせた数値。打者の総合的な能力を測る指標として用いられる。今回扱ったデータ上では、OPSの平均と得点数の相関係数は約0.957であり、非常に強い正の相関がある。

***2主成分分析**
多くの変数を持つデータをより少ない要素に集約する統計的手法

4.結果1

表1.主成分係数と分散の説明割合

		第一主成分	第二主成分	第三主成分	第四主成分	第五主成分	第六主成分
本塁打	第三四分位数	0.3683635	0.5258614	-0.5088995	-0.382408	-0.355084	-0.237448
	中央値	0.3522728	0.550805	0.518377	0.5231656	-0.009485	-0.173237
打率	第三四分位数	0.3768157	-0.42176	-0.484767	0.6393884	-0.170523	-0.085053
	中央値	0.4239529	-0.439055	0.3219289	-0.346709	0.1188611	-0.624116
OPS	第三四分位数	0.456296	0.1039727	-0.197269	-0.095147	0.7867303	0.3377442
	中央値	0.458563	-0.196526	0.3076638	-0.204901	-0.460081	0.6346506
	説明割合	0.6352517	0.1432494	0.0957303	0.0663336	0.0430785	0.0163566

上の6行は主成分係数を表しており、各数値とこの係数をかけて足し合わせると主成分の値になる。また、この係数からその主成分が何を表すのか分析できる。

第一主成分は、すべての項目の係数がほぼ等しいので、**すべての項目の平均の成績**を表していると考えられる。

第二主成分は、本塁打の2項目が大きい正の係数、打率の2項目が大きい負の係数、OPSは小さな値になっているため、**本塁打数が多いか、打率が高いか**という傾向を表す値だと考えられる。

第三主成分は、成績が第三四分位数以上の人数の3項目が負の値、成績が中央値以上の人数の3項目が正の値になっているため、**抜群の成績の選手の人数が比較的多いか、普通以上の成績の選手の人数が比較的多いか**を表す値だと考えられる。

第一主成分はほぼ明らかに得点と強い正の相関があるので、**第二、第三主成分を軸**にとって各データと各項目を**散布図**に表し、どのようなチームの得点数が多いかを調べる。

右の散布図において、各象限ごとに、原点に近いデータは除いて、**得点の平均**を表すと、表2のようになった。

表2.各象限ごとの得点の平均

第一象限	第二象限	第三象限	第四象限
601.4	544.3	578.5	602.3

打率の成績よりも本塁打数の成績が良いチームを**本塁打型のチーム**、本塁打よりも打率の成績が良いチームを**打率型のチーム**と呼ぶことにする。第1, 4象限は第2, 3象限に比べて得点の平均が大きいことから、第二主成分が正、つまり**本塁打型のチームのほうが打率型のチームよりも、得点数が多い傾向がある**と分かる。

第1象限と第4象限は得点の平均がほぼ同じだが、第2象限よりも第3象限の方が平均が高くなっている。**本塁打型のチーム**は、抜群の成績の選手の人数が多いか、普通以上の成績の選手の人数が多いかに**影響を受けにくい**が、**打率型のチームでは、抜群の成績の選手の人数が多いチームが、普通以上の成績の選手の人数が多いチームよりも、得点数が多い傾向にある**と分かる。

9.展望

・チームを大きく二つの型に分けて分析することができたが、**もう少し多くの型に分類**することができれば、もっと細かい分析ができると思うので、挑戦したい。
・打撃の成績だけから特徴を掴んだが、**盗塁**なども得点に影響を与える要素であると思ったので、それも加味して解析したい。
・攻撃の面からどのようなチームが得点を取りやすいのかを考えたが、今度は**投手成績**に着目して、どのようなチームが、チームとして失点が少なくなる傾向にあるのか調べたい。

5.方法2(重回帰分析)

次に、主成分分析によって得られた主成分が、具体的にはどのような割合で得点に
関与しているのかを調べるために、**第一から第三まで3つの主成分の値**を用いて**重回帰分析^{*3}**を行った。なお、同様に**標準化**後のデータを用いた。(①)

また、結果1で第二主成分の正負で第三主成分の得点への影響が大きく変わっていると予想されたため、**第二主成分が正の27個のデータと、第二主成分が負の21個のデータ**に分け、それぞれについて**標準化**した後、**重回帰分析**を行った。(②)

***3重回帰分析**

二つ以上の説明変数が目的変数に与える影響の度合いを調べる統計的手法

6.結果2

① **標準化された値として、得点をy、第一主成分をx₁、第二主成分をx₂、第三主成分をx₃とすると、**

$$y = 0.8116x_1 + 0.1739x_2 - 0.0674x_3$$

と表せると分かった。(図2)

この結果から、やはり**第一主成分の影響する割合が非常に大きく、第二主成分は大きい値であるほど得点が高くなっている**と分かり、**結果1と一致する**。

しかし、この方法では第三主成分がほぼ得点に影響しないことになってしまう。

② 一方、**第二主成分が正**であるデータだけで求めると、

$$y = 0.7109x_1 + 0.2701x_2 + 0.0183x_3$$

となり、**第二主成分が負**であるデータだけで求めると、

$$y = 0.8974x_1 - 0.0372x_2 - 0.0788x_3$$

となった。(図3、図4)

これより、第一主成分の影響する割合が大きいのは変わらないが、**第二主成分が正(本塁打型)であるデータは、第二主成分がかなり大きく影響し、第三主成分はほぼ完全に影響しないと分かる。第二主成分が負(打率型)であるデータは、逆に第二主成分があまり影響せず、第三主成分の方が大きく影響している**と分かった。また、第三主成分は値が小さいほど得点が高くなっていると分かり、これも**結果1と一致する**。

これらより、すべての項目の平均が最も重要であるが、本塁打型のチームでは、本塁打の成績がとて重要であり、打率型のチームでは、打率の成績はほとんど関係がなく、抜群の成績の選手の人数がかなり重要である。

※グラフについて

影響のとても小さかった成分(図4では第二主成分、その他は第三主成分)を無視し、各データを散布図にして、重回帰分析で求めた式を平面として表している。

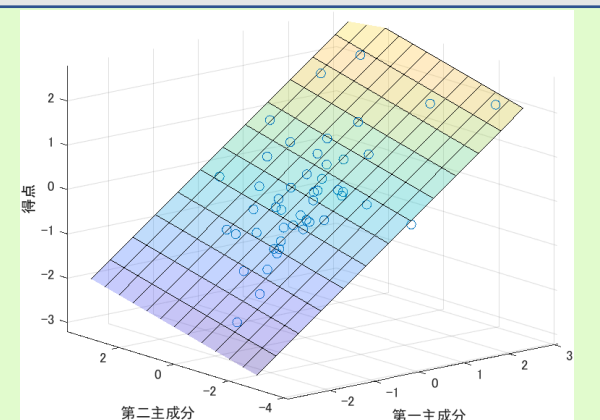


図2.第一主成分、第二主成分、得点を軸に取った散布図

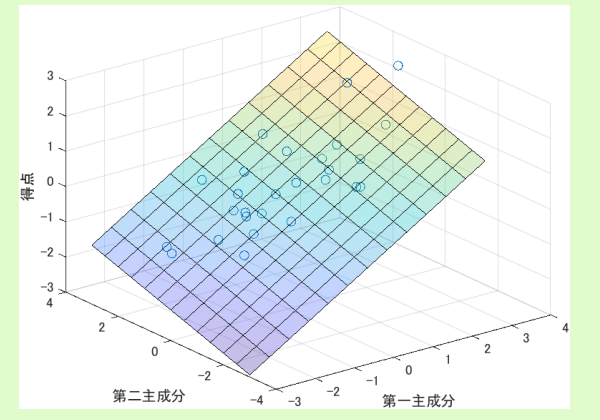


図3.第二主成分が正のデータで第一主成分、第二主成分、得点を軸に取った散布図

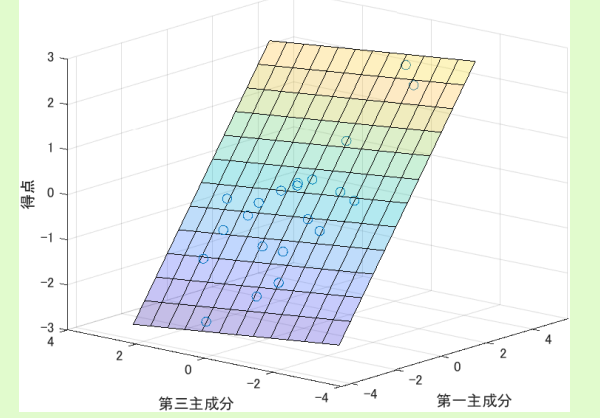


図4.第二主成分が負のデータで第一主成分、第二主成分、得点を軸に取った散布図

7.まとめ

- ・得点への影響の8割以上は全成績の平均が占めていてこれが最重要
- ・本塁打型：一般に打率型よりも得点が高い傾向で、本塁打数がとても重要だが、**抜群の成績の選手の数はほぼ関係がない。**
- ・打率型：一般に本塁打型よりも得点が高い傾向で、打率の成績は重要ではなく、**抜群の成績の選手の数がかなり重要である。**

8.考察

- ・本塁打型のチームのほうが、得点数が多い傾向があるのは、**本塁打は得点に直結する**のに対して、**打率は必ずしも得点にはつながらない**ことが影響すると考えられる。
- ・本塁打型のチームが、抜群の成績の選手の人数が多いか、普通以上の成績の選手の人数が多いかによる影響を受けにくいのは、**本塁打が多いチームはそもそも安定して得点が取れる**からだと考えられる。
- ・打率型のチームにおいて、抜群の成績の選手の人数が多いチームが、普通以上の成績の選手の人数が多いチームよりも、得点数が多いのは、打率が高くても、抜群の成績の選手が一定数いて、**ランナーを帰さない**と、得点を取ることができないからだと考えられる。
- ・これらの結果から、得点力の高いチームを作るには、全体的な成績を上げるのはもちろんだが、できるなら**本塁打型のチームを目指す**べきである。しかし、もし打率型で得点力の高いチームを目指すのなら、**抜群の成績の選手の数を増やす**ようにすべきである。

10.謝辞

本研究のためのデータ提供には、『情報・システム研究機構 統計数理研究所 医療健康データ科学研究センター』および『データスタジアム 株式会社』の皆様にご支援頂きました。心から感謝申し上げます。また、本研究を行う機会を用意して下さった『日本統計学会 スポーツ統計分科会』および『スポーツデータ解析コンペティション』に携わられている皆様にもお礼申し上げます。