

過去15年間のプロ野球で、最も効率的に戦ったチームはどこか？

研究動機・目的

現在存在する野球の指標のほとんどは「いかに良い成績を残したか」を表す指標である。しかし対照的に「いかに効率よく戦ったか」を表す基準・指標は少なく、とりわけチーム全体に関してこの類の指標はないと言っても過言ではない。

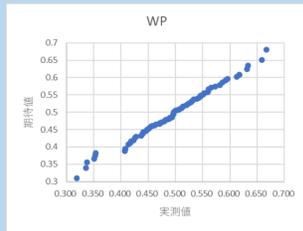
そこで、チームを対象とした「**戦った効率**」を測る指標を提案すること、そして実際にその指標を用いて問いに対する答えを導き出すことが目的である。今回は主に**戦力面・資金面と勝率の比較**による分析を行った。

基本的な前提・確認事項

野球は「**打撃・走塁・投球・守備**」の4つの観点に分けることができる。これは野球のすべてのプレーがこれらによって分類できるためである。実際にWARという指標では、個人の貢献度をこれら4つに分けて数値化している。そのため、右の図のような分類が可能であるとして、本研究を進めた。



また、各分析で必要に応じて標準化・偏差値の算出を行っているが、あらかじめ右のような正規確率プロットを用いて正規性を確認している。そのため、標準化・偏差値の算出については、**正規性を確認した上での分析である**ことを、ここで断りを入れさせて頂く。

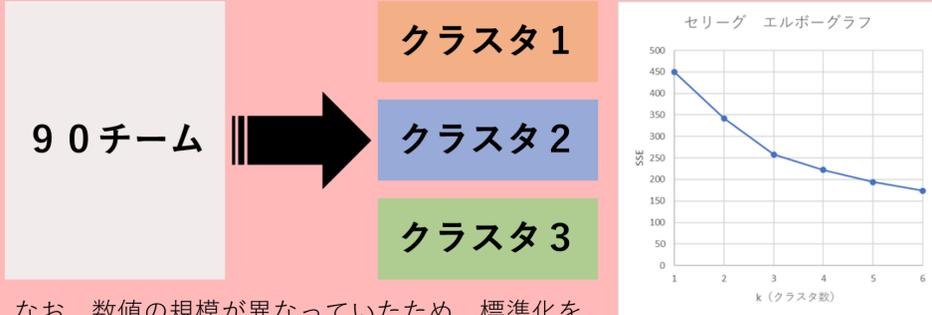


分析対象は2005～2019年まで15年間の両リーグ各6球団のデータ、計90チーム分の成績・資金のデータである。

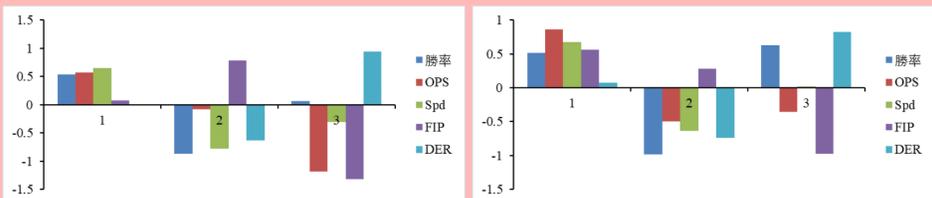
分析 i クラスタリング；チームを分類する

チームにはそれぞれ色と言われるものがある。攻撃型や守備型、機動力型などというものだ。それらの型によって**重要な力も変わってくる**。打撃重視のチームにとって中途半端な機動力は意味が薄いし、逆に打撃が弱いチームにとっては得点するのに必須の力となる。

この違いを考慮するべく、分析する対象に**クラスタリング**を行った。使用したデータは5つ。チームの強さを表す勝率(WP)と、前述した**4観点の力を測る OPS・Spd・FIP・DERという指標**である。方法は非階層クラスタリングの1つである**k-means法**を選択し、**エルボー法**によってクラスタ数は3つに決定した。



なお、数値の規模が異なっていたため、標準化をしたうえで分析した。



各クラスタの特徴（左：セリーグ 右：パリーグ）

データ元 「日本野球機構」 <https://npb.jp>、「1.02 - Essence of Baseball」 <https://1point02.jp/op/index.aspx>、「日本プロ野球選手会 公式ホームページ」 <http://jpbpa.net/>
参考文献 清水裕士 フリーの統計分析ソフトHAD：機能の紹介と統計学習・教育、研究実践における利用方法の提案、(2016)

分析 ii 重回帰分析；戦力の効率を測る

分析 i のクラスタごとに、**目的変数を勝率、説明変数をOPS・Spd・FIP・DER**とした**重回帰分析**を行った。この分析の意図は、チームの基本的な能力から勝率を予測する式を立てることである。勝率はシンプルにチームの強さを表すため、能力から予測した勝率はすなわち「**チームの潜在的な戦力**」を表していると考えた。いわゆるポテンシャルである。これと実際の勝率を比較することで、そのチームが戦力を使いこなせたのかを測る指標が定義できる。



まず、重回帰分析によって以下の回帰係数が得られた。右の表では、決定係数などの主な数値をクラスタごとに記載している。（上：セリーグ 下：パリーグ）

	1	2	3
切片 :a	-0.477	-2.070	-2.608
OPS :b	1.130	1.823	2.048
Spd :c	-0.011	0.017	-0.032
FIP :d	-0.115	-0.125	-0.073
DER :e	0.973	2.455	3.019

	1	2	3
切片 :a	-0.890	-0.759	-0.973
OPS :b	0.888	1.076	1.323
Spd :c	0.030	0.030	0.015
FIP :d	-0.043	-0.075	-0.061
DER :e	1.151	0.904	1.048

	1	2	3
重相関 R	0.842	0.895	0.873
重決定 R2	0.710	0.801	0.762
補正 R2	0.679	0.766	0.694

	1	2	3
重相関 R	0.764	0.831	0.854
重決定 R2	0.583	0.691	0.730
補正 R2	0.517	0.647	0.681

こうして求めた各回帰式で求まる値を、**予測勝率(eWP)として定義**する。なお、a～eは表の文字のことを指し、係数は所属リーグ・クラスタによって変動する。

$$eWP = a + b \times OPS + c \times Spd + d \times FIP + e \times DER$$

これと勝率(WP)を比較したものが、戦力の効率を表す指標になる。ここでは**戦力効率(WE)**と呼び、以下のように定義する。

$$WE = \frac{WP}{eWP}$$

この戦力効率は、値が1ならば戦力に見合った成績であることを指し、大きければ効率的、小さければ非効率的であることを表す。

補足 クラスタリングの指標化

戦力効率を求めるためにはチームがどのクラスタに所属しているのかを知る必要がある。しかし新規データに指標を使用する際、再びクラスタリングを行う必要はない。**新規データとクラスタの重心との距離**を求めれば良いからだ。

$$距離 = \sqrt{(WP' - \overline{WP})^2 + (OPS' - \overline{OPS})^2 + (Spd' - \overline{Spd})^2 + (FIP' - \overline{FIP})^2 + (DER' - \overline{DER})^2}$$

「'」が付いた部分には新規データの標準化したものを入れ、「-」が上についたものは各重心の対応する値を入れる。そして3種類出てくる数値の内、最小になる場合のクラスタ番号を選択する。

計算に必要な重心は右の表の通り。ただし、データは標準化されているので、新規データも標準化する必要がある。よってその下の表にある平均値・標準偏差を定数として扱い、新規データを標準化する。（左：セリーグ 右：パリーグ）

分析 iii 偏差値の導入；資金の効率を測る

ここではチームが持つ資金の大きさとチームの成績について考える。まず前提として、少ない資金でより良い成績を残せることが優秀であるという、一般的な**コストパフォーマンスの発想で指標を定義**する。分析 ii で戦力効率を「**値が1ならば戦力に見合った成績であることを指し、大きければ効率的、小さければ非効率的であることを表す。**」としたため、資金効率も同様の特性を持つようにする。



チームの資金には経費など色々なものがあるが、ここではシンプルに選手に支払われる年俵で測ることとした。使用するデータは合計年俵ではなく**平均年俵**であり、これはプロ野球選手会の情報では公示されていない外国人選手の年俵による影響を抑えるためである。

ここでは**平均年俵=資金力(FP)**（ただし年俵の単位は万円）として考える。**資金効率(FE)を定義**するため、戦力効率と同様に資金力で勝率を割る形の指標を仮に定義した。しかしこれでは値の規模が異なりすぎていて参考にならない。そこで両データの**偏差値を利用**することで、戦力効率と同じような「共に同じ水準なら1になる」という指標にすることができた。

$$FE = \frac{WP''}{FP''}$$

「''」は偏差値であることを示している。標準化ではなく偏差値を用いたのは、標準化では平均が0になってしまうため、分母が0になったり割合でありながら負の数になってしまうからである。その点、**偏差値は基本的に正の数の範囲で計算ができるため、適当であると判断した**。ちなみに、偏差値の計算に必要な平均値と標準偏差は右の表の通り。（上：セリーグ 下：パリーグ）

	WP	FP
平均値	0.495	3935
標準偏差	0.073	1232

	WP	FP
平均値	0.505	3642
標準偏差	0.071	1038

結論・考察 総合効率の定義・問いの答え

分析 ii・iiiで定義した2つの効率は、共に同じ性質を持つようにしてある。そのためこれらを合算すれば、同じ尺度で総合的な効率を測る「**総合効率(CE)**」を定義できる。

$$CE = WE \times FE$$

$$(WE = \frac{WP}{eWP} \quad FE = \frac{WP''}{FP''})$$

この総合効率は戦力効率・資金効率と同様、値が1ならばチームに見合った成績であることを指し、大きければ効率的、小さければ非効率的であることを表す。実際にこの総合効率を用いて過去15年間の各球団を分析すると、最も総合効率が高かったのは

- ✓ **セリーグ：2017年 広島東洋カープ** (CE=1.759)
- ✓ **パリーグ：2007年 北海道日本ハムファイターズ** (CE=1.605)

だった。どちらも通常の同レベルチームより1.5倍以上勝っている計算になる。そしてどちらもこの年に優勝をしている。つまり、効率的に戦うことは**優勝をする為に必要なことであるとも言える**のではないかと、とも考察できた。

	WP	OPS	Spd	FIP	DER
1	0.544	0.574	0.648	0.075	0.000
2	-0.878	-0.077	-0.787	0.783	-0.640
3	0.061	-1.186	-0.308	-1.325	-0.944

	WP	OPS	Spd	FIP	DER
平均値	0.495	0.704	4.06	3.68	0.688
標準偏差	0.073	0.040	0.49	0.47	0.012

謝辞 このような貴重な機会を設けてくださった方々、ご支援いただいた『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』様に、心より感謝申し上げます。