

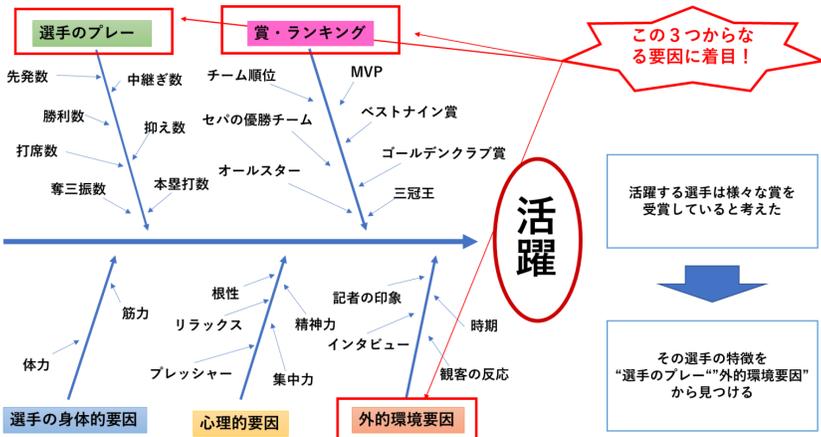
活躍とは何か？人の心を測る指標の作成

～プロ野球投手データから、「記者」の心を数値化する～

研究動機

大谷選手のMVPの獲得について、今年も全米や日本の大きな話題となった。そこで感じることは、「スポーツにおいて活躍とは何か」ということだ。それを定義し、数値化できないか。人の五感を数値化する先行研究はあるが、スポーツの活躍を研究した先行例は見当たらなかった。そこで、その先陣を切ろうと、プロ野球の投手データを用いて、研究を進めることにした。プロ野球において、MVPやベストナイン、オールスターの投票データ量は予想よりはるかに少なく、人の心の分析に使用できると判断できたのはMVPの投票データだけであった。MVPは、スポーツ担当記者1人につき、1位（5点）、2位（3点）、3位（1点）の3票が与えられ、その投票された点数の合計が最高となった選手がMVPとして選出される。そこで2017年度～2021年度の5年間のMVP投票結果を収集して、記者の心の数値化を目指して研究を開始した。

特性要因図



分析①

仮説 2017年～2021年までのMVPの得票を得た投手72人をピックアップし、そのMVPの点数（1位5点、2位3点、3位1点の合計）と提供していただいた投手成績の全項目との相関係数を調べ、どの項目と相関があるかを調査することにした。そして、強い相関が現れる項目は必ずあると予想した。

分析結果: 相関係数の上位5項目は表1の通りである。0.5未満の値ばかりで、強い相関の項目は、見られなかった。

	勝利	完封勝利	奪三振	完投	投球回
相関係数	0.454548	0.424446	0.38777	0.360719	0.338333

分析②

仮説 MVPの点数は、票の獲得順位が上がるごとに、指数関数的に増加していると感じた。そこで対数を取り、1次関数や2次関数などの整関数の状態に変換したいと考えた。常用対数 \log_{10} (MVP点数)を用いて再度相関係数を調査した。

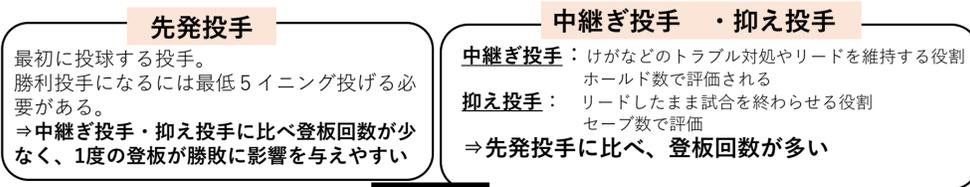
分析結果: 相関係数の上位5項目は、表2のようになり、明らかに係数は増加した。また、対数の底を変えても相関係数は変化しなかった。このことから以下の2点を考察できた。

	勝利	奪三振	投球回	完封勝利	投球数
相関係数	0.577688	0.553773	0.497654	0.491911	0.479365

・ウェーバー・フェヒナーの法則
「人間の感覚の大きさは、受ける刺激の強さの対数に比例する」という法則が提唱されていて、MVP得票にも当てはまることがわかった。

・対数の底を変えても相関係数が変化しない理由の考察
2つの偏差ベクトルを $\vec{x} = (x_1 - \bar{x}, \dots, x_n - \bar{x})$, $\vec{y} = (y_1 - \bar{y}, \dots, y_n - \bar{y})$ としたとき
相関係数 $r = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| |\vec{y}|}$ であり、 $X = ax + b (a > 0)$ と変換しても、Xとyの相関係数はrである。
ここで、 $\log_p x = \frac{\log_q x}{\log_q p}$ であるから、相関係数は変化しない。

分析③



考察 X=勝利数、Y=セーブ数、Z=ホールド数として、活躍を示す指標の形を下のように設定して、a～fの6つ定数を決定したいと考えた。

$$\text{新指標の形} = \begin{cases} \text{先発投手のとき: } aX + bY + cZ \\ \text{中継ぎ・抑え投手のとき: } dX + eY + fZ \end{cases}$$

(1)一試合で先発投手が5イニング以上投げるのに対し、中継ぎ・抑え投手は大体1~2イニング投げることから先発投手の勝利数を約4倍にするとよいと予測した。そこで、まず、 $a=4, b=c=d=e=f=1$ とおいてみると、相関係数0.71という強い相関がみられた。a以外の係数も変化させて、さらに高い相関が得ようと考えた。
(2)MVP点数の2乗根、3乗根、4乗根を取り、対数以外にも、決定された指標と強い相関を示す関数があるかもしれないと考え、調査してみた。

分析結果

(1) $b=1$ を固定してa, c～fの値を様々に変化させると、表3の値のときに相関係数の最大値0.7236を得た。
(2)それぞれの関数における相関係数は表4のようになり、対数を用いた場合が最も高い結果となった。心と対数の関係の深さを強く感じた。

a	b	c	d	e	f
4.16	1	3.79	0.90	0.93	1.01

	対数	2乗根	3乗根	4乗根
相関係数	0.723629	0.667454	0.702489	0.715115

(1)(2)より、X=勝利数、Y=セーブ数、Z=ホールド数として、以下の指標①が得られた。

$$\text{指標①} = \begin{cases} \text{先発投手のとき: } 4.2X + Y + 3.8Z \\ \text{中継ぎ・抑え投手のとき: } 0.9X + 0.9Y + Z \end{cases}$$

分析④

仮説 MVPは年度末に選考されることから選手の前期成績と後期成績で、後期成績のほうが記憶に新しく、より影響力があると考えた。そこで、指標①で一年間の成績にかけていた係数を、前期成績と後期成績それぞれ異なる係数をかけ、対数値との相関係数を調べることにした。

先発投手	前期X	a1前	前期Y	b1前	前期Z	c1前	後期X	a1後	後期Y	b1後	後期Z	c1後
中継ぎ・抑え投手	前期X	a2前	前期Y	b2前	前期Z	c2前	後期X	a2後	後期Y	b2後	後期Z	c2後
相関係数	-0.637	4.31	1	5.21	2.1	1	-4.58	0.999	0.0169	5.45	-0.511	0.285

分析結果 相関係数は、指標①の時より下がり、あまり相関が見られなくなった。このことから、記者が投票する際、前期と後期の成績の影響に差異はなく、一年間の成績を通して判断していると考えられる。

	指標①	指標(前期後期別)
相関係数	0.7235	0.544615

分析⑤

仮説 指標①を「シーズン個人投手成績」を用い、降順に並べ替えると、MVPを獲得した選手が上位に並び予想していたが、MVPを獲得した選手間にMVPを獲得していない選手がたくさんいた。そこでMVP選手がもっと上位に並び新しい指標を作ろうとした。

新指標作成方法

選手が所属するチームの順位が高いほどMVPの得票は、得やすいのではないかと考えた。そこで1位、2位のチームの選手の指標①にどのような定数をかけるべきか調査し、MVP点数対数値との相関係数の最大値は表6のようになり、指標②を決定した。

	0.8	0.9	1
1位	0.740165	0.739056	0.723629
2位	0.746653	0.751552	0.744397
1.2	0.738866	0.747162	0.745946

指標② = $\begin{cases} \text{1位チームの選手の時: } \text{指標①} \times 1.1 \\ \text{2位チームの選手の時: } \text{指標①} \times 0.92 \\ \text{他のチームの選手の時: } \text{指標①} \end{cases}$

目的変数: MVP点数の対数としての単回帰分析において補正R2値0.584 p-値 3.29×10^{-15}

検証

2017～2021年度の5年間のべ1714選手を、年度ごとに指標①、指標②それぞれ降順に並び替えて、MVPの得票を1票でも得た全72人の選手の順位のとると、指標①よりも指標②の和の方が74減少した。

指標①	順位	和
645	74	571

考察 2位のチームの選手に0.9をかけることは大きな驚きだ。この分析からは、2位のチーム選手は、3位以下のチーム選手より少し厳しい感情を加えられる結果となった。

分析⑥

分析・結果 指標②を作成したが、それを支えてきたデータであるMVP点数は、1位5点、2位3点、3位1点の合計である。この重みの付け方は、本当に活躍を表すに相応しいか検証するため、逆に、指標②を固定し、1位5点、2位p点、3位q点として、最も相関が高まるp,qの値を調査してみた。

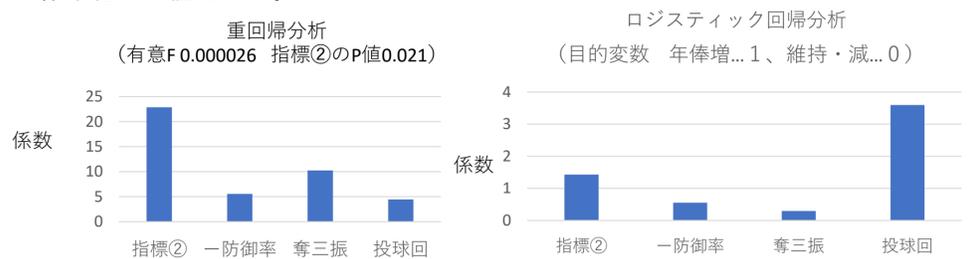
$p=6.8 (\approx 5), q=5.2 (\approx 5)$ の時に相関係数の最大値0.768を得た。

提言

指標②を信じるならば、年間を通した活躍を表すMVPの決定指標としては、1位～3位の重みを変えず、記者が1人3票ずつ投票した得票数の合計の方が、より相応しいと言えるのではないかと。

分析⑦

指標②は、活躍を表す数値の1つである年俸の増減とどのように相関があるかを重回帰分析とロジスティック回帰分析を用いて検証してみた。212人の投手について、2021年度と2020年度の「年俸」の増減を目的変数とし、説明変数は「指標②」「-防御率」「奪三振」「投球回」の各2021年度と2020年度の増減を標準化した値とした。



年俸は、人の心を素直に表しているとは言えないため、十分な結果は得られないと予想していたが、重回帰分析において、指標②は、他の項目よりかなり大きな係数値が示された。

総括

・ウェーバー・フェヒナーの法則はMVPの投票数にも当てはまり、対数値は人の心の数値化に適することを示す例となった。
・指標②は、目的であった記者の心を数値化して、投手の活躍を評価する指標と言え、MVP獲得や年俸増減の予想も十分に行える指標となりそうだ。
・MVPの投票は、1位～3位に点数差をつけず、票数のみで決定する方法も検証すべきだ。

今後の展望

・奪三振、防御率などの項目を加えたり、クラスター分析を用いて活躍する選手を特徴に応じて分類したりして、より精度の高い指標を作成したい。
・打者の項目を調査し、打者の活躍を測る指標を作成したい。
・完成した指標を用いて来年度のMVP獲得者を予想し、その精度を確認したい。

謝辞

このような機会を与えてくださりデータを提供してくださった情報・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の皆様、データスタジアム株式会社の皆様、並びに研究に必要な分析手法やプログラミングについてご教授くださった徳島文理大学 理工学部 電子情報工学科 山本由和教授、本校先生方にお礼申し上げます。

参考文献・データセット

- 【参考文献】
 - 空間情報クラブ(ウェーバー・フェヒナーの法則人間の五感を数値化)
 - EXCEL ビジネス統計分析【ビジテク】第3版
- 【データセット】
 - SPAIIA スポーツ×AI×データ解析総合メディア
 - Wikipedia
 - nf3-Baseball Data House-月間成績
 - NPB.jp 日本野球機構