

プロ野球における年間勝率と順位決定要素

芝浦工業大学 数理科学研究会 BV17057 西脇 友哉

平成 29 年 11 月 3 日

研究動機

私は 2 年前に“MLB と NPB での QS (クオリティ・スタート) が持つ価値の違い”について研究していた。その際に「多くの試合で 6 回までを無失点に抑えられるものの勝率が悪い球団」「6 回までを無失点に抑えたときの勝率は高いがそういった試合が少ない球団」などが存在することを発見した。そこで、6 回 n 失点 ($n = 0, 1, 2, \dots$) という局面について、各 n に対し試合数と勝率という 2 項目での評価を基にして考察を図った。

1 当研究の方向性

過去の成績 ('14~'17) を元に、試合比率 (年間の公式戦試合数¹に対して 6 回 n 失点の試合数が占める割合) と状況別勝率 ('6 回 n 失点のとき) という特定条件下での勝率) の考察を各 n について行うことで、2018 年シーズン以降各球団がリーグ優勝を目指すにあたって達成すべき基準を定める。

2 優勝するために必要な勝率

優勝ラインを設定するために、'07~'17 の 11 年間を対象に優勝チームのグループ、2 位チームのグループをそれぞれ作成し、マハラノビス距離による判別分析を行った。

年間勝率について、優勝球団 (1 位) は $\mu_1 = 0.596$, $\sigma_1 = 0.043$, 2 位の球団は $\mu_2 = 0.548$, $\sigma_2 = 0.024$ であった。²ここで、

$$D_1 = \frac{\mu_1 - x}{\sigma_1}, \quad D_2 = \frac{x - \mu_2}{\sigma_2}$$

とすると、優勝ラインは $D_1 = D_2$ となる年間勝率より求められるので、これを x について解くと、

$$x = \frac{\sigma_2\mu_1 + \sigma_1\mu_2}{\sigma_1 + \sigma_2} = 0.565$$

を得る。これを優勝ラインとして定める。そして、この数値に最も近い勝敗成績が (78-60-5) であったことから、これを目安に n ごとの必要勝利数を導出していく。

ちなみに '14 ソフトバンク (1 位) の成績が (78-60-6) であるため、勝敗の分布に対するイメージを表 1 から膨らませることはできるが、6 回の失点状況の基準を設定する際には '14 ソフトバンクのデータをそのまま採用するのではなく、改めて計算する必要がある。

表 1: '14 年度ソフトバンクの勝敗成績

n	勝	敗	分	試合数	試合比率	状況別勝率
0	20	3	2	25	0.174	0.870
1	20	3	1	24	0.167	0.870
2	19	11	2	32	0.222	0.633
3	11	12	1	24	0.167	0.478
4 以上	8	31	0	39	0.271	0.205
計	78	60	6	144	1	年間勝率 0.565

3 年間勝率を重回帰式で表現する

$n = 0, 1, 2, 3$ のとき、また $n \geq 4$ のときに分類し、 n を固定する。その後、試合比率 (x_1) と状況別勝率 (x_2) を説明変数、年間勝率 (Y) を目的変数として重回帰分析を行う³。標本分散共分散行列と重回帰式による分散共分散行列との比較により、

$$Y_0 = a_1x_1 + a_2x_2 + a_0$$

$$Y_1 = b_1x_1 + b_2x_2 + b_0$$

⋮

$$Y_n = a'_1x_1 + a'_2x_2 + a'_0$$

という形式で重回帰式が得られる⁴(詳細は資料にて掲載)。

ただし、得られた重回帰式だけでは変数が 2 つ残っている状態なので、2 つの説明変数のうちのいずれかを固定しなければ全ての数値を求めることは出来ない。状況別勝率には打線の援護や救援投手の出来など複数の要因が絡むのに対し、試合比率は決定要素のほとんどが先発投手の出来による。よって今回は複雑な変数を先に処理するために状況別勝率を固定した。

4 CS 出場ライン、最下位回避ライン

“2 優勝するために必要な勝率” で定義した優勝ラインの他に、一次元のマハラノビス距離による判別分析の手法により、CS 出場ライン (3 位) や最下位回避ライン (5 位) についても設定することができる。ここでは途中計算を省略するが、 $\mu_m, \sigma_m (m = 3, 4, 5, 6)$ を用いて CS 出場ライン: 0.505, 最下位回避ライン: 0.433 を得た。それらについても優勝ラインと同様に重回帰分析を行うことが出来る。

今後の課題

元々 QS は投手個人の記録という側面があるが、今回はチームの失点を対象に資料を収集したため、先発投手一人一人の成績まで踏み込んだ考察を行うことが出来なかった。今後選手個人に着目していく場合は、QS の達成条件が「6 回以上かつ自責点 3 以内」であることに注意し、失点と自責点の違いを正しく処理する方法を考えなければならない。

参考文献

- [1] NPB.jp 日本野球機構, <http://npb.jp/>, 最終アクセス: 2017.10.27.
- [2] 森田 浩 著, 図解入門ビジネス 多変量解析の基本と実践がよ〜くわかる本, 秀和システム, 2014.
- [3] 室 淳子, 石村貞夫 著, Excel でやさしく学ぶ多変量解析 [第 2 版], 東京図書株式会社, 2007.

¹'14:144 試合, '15~:143 試合

² x :年間勝率 (変数), μ :平均, σ :標準偏差, D_1, D_2 :1 位群, 2 位群それぞれからのマハラノビス距離

³この際、試合比率と状況別勝率との間に多重共線性が無いことを確認する。

⁴ x_1 と x_2 の係数は全て偏帰係数。