

セイバーメトリクスの妥当性についての研究

2007MI004 秋江一希

指導教員：松田眞一

1 はじめに

セイバーメトリクスの指標では、「打率」、「防御率」、「打点」といった従来の指標だけではわかりにくかった守備の力や個人の力でどれだけ得点を生み出すことができるかなどがわかるようになった。そして近年、セイバーメトリクスの新しい指標がどんどん作られている。セイバーメトリクスが普及しているメジャーリーグと日本のプロ野球で何か違いはあるのか、そしてどんどん作られているセイバーメトリクスの指標の中で、改良する余地のある指標はないのかと思い、セイバーメトリクスの分析を卒業研究のテーマとして取り組んだ。

2 セイバーメトリクス

野球において、選手の評価方法などを統計データを分析して作られた様々な指標の総称をセイバーメトリクスという。(J. アルバート [1]、高橋 [3] 参照)

3 データについて

データは、Web[5] のホームページから 2001 年度から 2010 年度までのメジャーリーグ全 30 球団の「勝率」、「 BB/K 」、「 OPS 」、「 ISO 」、「 $BABIP$ 」、「 WRC 」、「 $K/9$ 」、「 $BB/9$ 」、「 K/BB 」、「 $HR/9$ 」、「 $WHIP$ 」、「被 $BABIP$ 」、「 FIP 」の 14 項目と「 XR 」を求めするために 1998 年度から 2010 年度の「単打」、「二塁打」、「三塁打」、「本塁打」、「四球」、「死球」、「故意四球」、「盗塁」、「盗塁刺」、「打数」、「安打」、「三振」、「併殺打」、「犠飛」、「犠打」、「得点」の 16 項目を用いた。日本プロ野球のデータは、Web[4] のホームページから 2010 年度の日本プロ野球全 12 球団の「単打」、「二塁打」、「三塁打」、「本塁打」、「四球」、「死球」、「故意四球」、「盗塁」、「盗塁刺」、「打数」、「安打」、「三振」、「併殺打」、「犠飛」、「犠打」、「得点」の 16 項目を用いた。

4 解析方法

「勝率」を目的変数とし、打者指標「 BB/K 」、「 OPS 」、「 ISO 」、「 $BABIP$ 」、「 WRC 」と投手指標「 $K/9$ 」、「 $BB/9$ 」、「 K/BB 」、「 $HR/9$ 」、「 $WHIP$ 」、「被 $BABIP$ 」、「 FIP 」に分け、それぞれを説明変数として重回帰分析した。なお、変数選択は、修正決定係数による減少法を用いた。

5 重回帰分析の結果

5.1 2001 年～2010 年

ここでは、2010 年の分析結果を示す。まず、全 30 球団の打者指標のデータで重回帰分析した。すると OPS 、 ISO が残り、決定係数 0.5939、修正決定係数 0.4490 で、全 30 球団の投手指標では、 $K/9$ 、 $BB/9$ 、 K/BB 、 $WHIP$ が残り、決定係数 0.712、修正決定係数

0.6645 である。打者指標、投手指標ともに勝率をある程度は説明できる。次に、指名打者制の影響を調べるために、アメリカンリーグの球団とナショナルリーグの球団に分けて、打者指標と投手指標それぞれを目的変数として重回帰分析をした。アメリカンリーグの打者指標では、 BB/K 、 $BABIP$ 、 WRC が残り、決定係数 0.7977、修正決定係数 0.7351 で、投手指標では、 $BB/9$ 、 K/BB 、 $HR/9$ 、 $WHIP$ 、 FIP が残り、決定係数 0.9127、修正決定係数 0.8571 である。ナショナルリーグの打者指標では、 OPS 、 ISO が残り、決定係数 0.5568、修正決定係数 0.4490 で、投手指標では、 K/BB 、 $WHIP$ が残り、決定係数 0.8637、修正決定係数 0.7826 である。この結果より指名打者制は、勝率に影響を与えている。

5.2 攻撃重視のチーム (ヤンキース)

攻撃重視のチームとして Web[5] から 2001 年から 2010 年のヤンキースの打者指標のデータで重回帰分析した。すると BB/K 、 ISO 、 $BABIP$ 、 WRC が残り、決定係数 0.8852、修正決定係数 0.7881 で、投手指標では、 $K/9$ 、 $BB/9$ 、 K/BB 、 $HR/9$ が残り、決定係数 0.7891、修正決定係数 0.3210 である。したがって、攻撃重視のチームは、減少法によって残った打者指標 BB/K 、 ISO 、 $BABIP$ 、 WRC 、投手指標 $K/9$ 、 $BB/9$ 、 K/BB 、 $HR/9$ と勝率は妥当性があると考えられる。

5.3 守備重視のチーム (ブレーブス)

守備重視のチームとして Web[5] から 2001 年から 2010 年のブレーブスの打者指標のデータで重回帰分析した。すると $BABIP$ が残り、決定係数 0.3919、修正決定係数 0.0994 で、投手指標では、 $K/9$ 、 $WHIP$ が残り、決定係数 0.7704、修正決定係数 0.5927 である。したがって、守備重視のチームでは、減少法によって残った投手指標 $K/9$ 、 $WHIP$ と勝率は妥当性があると考えられる。

5.4 投打のバランスの良いチーム (フィリーズ)

投打のバランスのよいチームとして Web[5] から 2001 年から 2010 年のフィリーズの打者指標のデータで重回帰分析した。すると OPS が残り、決定係数 0.4088、修正決定係数 0.0418 で、投手指標では、 $K/9$ 、 $BB/9$ 、 $HR/9$ 、 FIP が残り、決定係数 0.9178、修正決定係数 0.8093 である。したがって、投打のバランスのよいチームでは、減少法によって残った投手指標 $K/9$ 、 $BB/9$ 、 $HR/9$ 、 FIP と勝率は妥当性があると考えられる。

5.5 まとめ

攻撃重視のチーム、投打のバランスのよいチーム、守備重視のチームとでわけた分析結果をみると、どの場合も投手指標では、高い決定係数が得られた。これより野球は、守備が重要である。

6 XR について

XRとは、RCと同様個人の得点創出能力を表す総合指標の一つである。(Web[2] 参照)

$$XR = 0.50 \times \text{単打} + 0.72 \times \text{二塁打} + 1.04 \times \text{三塁打} + 1.44 \times \text{本塁打} + 0.34 \times (\text{四球} + \text{死球} - \text{故意四球}) + 0.18 \times \text{盗塁} - 0.32 \times \text{盗塁刺} - 0.090 \times (\text{打数} - \text{安打} - \text{三振}) - 0.098 \times \text{三振} - 0.37 \times \text{併殺打} + 0.37 \times \text{犠飛} + 0.04 \times \text{犠打}$$

Web[2]よりXRは、1955年から1997年のメジャーリーグのチーム成績を元に重回帰分析を行い修正を加えて式を決定したものである。近年のメジャーリーグの傾向から値が変わってきていると考えメジャーリーグが30球団になった1998年から2010年の13年間のチーム成績からXRを求める。

表1を元にXRを作成する。

表1 1998年から2010年(目的変数:得点)の重回帰分析の結果

項目	回帰係数	p 値
切片	234.8824	0.0986
単打	0.5483	1.5×10^{-74}
二塁打	0.6997	2.8×10^{-39}
三塁打	1.1304	6.2×10^{-16}
本塁打	1.4066	2.1×10^{-121}
(四球+死球-故意四球)	0.3666	3.5×10^{-56}
故意四球	0.0451	0.6055
盗塁	0.0714	0.1305
盗塁刺	-0.2194	0.0709
(打数-安打-三振)	-0.1523	2.3×10^{-6}
三振	-0.1845	3.0×10^{-8}
併殺打	-0.4039	6.2×10^{-5}
犠飛	0.4849	0.0026
犠打	-0.0752	0.2710

$$XR' = 0.55 \times \text{単打} + 0.70 \times \text{二塁打} + 1.13 \times \text{三塁打} + 1.41 \times \text{本塁打} + 0.37 \times (\text{四球} + \text{死球} - \text{故意四球}) + 0.07 \times \text{盗塁} - 0.22 \times \text{盗塁刺} - 0.15 \times (\text{打数} - \text{安打} - \text{三振}) - 0.18 \times \text{三振} - 0.40 \times \text{併殺打} + 0.48 \times \text{犠飛} - 0.08 \times \text{犠打}$$

となる。t 統計量の有意確率(p 値)が0.1より小さければ傾向と評価する。傾向と評価できるのは、故意四球、盗塁、犠打以外である。変化が大きかった(打数-安打-三振)は-0.090から-0.15、単打は、0.50から0.55に回帰係数が変わった。これは、進塁打の得点への影響が下がり、単打が上がった。Web[2]よりメジャーリーグの平均球速は、2002年の約89.9マイルから2010年では、91.2マイルとなり、パワーピッチャーが増加したことが単打に影響していると考えられる。

XR'は、日本プロ野球のチームの得点と大きくずれたので、(打数-安打-三振)は既存のXRの値を使う。したがって、

$$XR'' = 0.55 \times \text{単打} + 0.70 \times \text{二塁打} + 1.13 \times \text{三塁打} + 1.41 \times \text{本塁打} + 0.37 \times (\text{四球} + \text{死球} - \text{故意四球})$$

$$) + 0.05 \times \text{故意四球} + 0.07 \times \text{盗塁} - 0.22 \times \text{盗塁刺} - 0.09 \times (\text{打数} - \text{安打} - \text{三振}) - 0.18 \times \text{三振} - 0.40 \times \text{併殺打} + 0.48 \times \text{犠飛} - 0.08 \times \text{犠打}$$

となる。ここで、改良した指標XR''と既存の指標XRを2010年の日本プロ野球全12球団のデータで比較する。

表2 2010年の日本プロ野球のデータで指標の比較

チーム	得点	XR	XR''
阪神	740	755.91	713.26
ヤクルト	617	672.84	627.32
巨人	711	746.11	690.36
広島	596	613.03	563.95
中日	539	604.86	545.08
横浜	521	562.85	503.68
ロッテ	708	711.39	661.18
日本ハム	612	625.21	575.38
西武	680	716.39	663.29
オリックス	644	682.88	633.80
ソフトバンク	638	646.59	586.77
楽天	576	608.32	564.38

6.1 まとめ

チームの成績からXRを求めると値がチームの得点に近づくことが指標として妥当性があるといえる。ヤクルト、巨人、中日、横浜、西武、オリックス、楽天の7球団でXR''のほうがよい値が出た。また得点との相関係数は、XRは0.9626でXR''では0.9621とほぼ同じ高い値より日本プロ野球においてXR''のほうが指標として妥当性があるといえる。したがって、日米で進塁打が得点に与える影響は違う。

7 おわり

今回の研究で野球についてセイバーメトリクスを使うことでいろいろな視点から見ることができ、これからの野球を見るにおいて楽しみが出来た。これからもセイバーメトリクスの指標は作られると思うので、その指標の妥当性について興味を持った。

参考文献

- [1] J. アルバート/J. ベネット 著, 後藤寿彦 監修, 加藤貴昭 訳: 『メジャーリーグの数理科学 上・下』, シュプリンガー・フェアラーク東京, 2004.
- [2] XR について
<http://www.baseballthinkfactory.org/btf/scholars/furtado/articles/IntroducingXR.html>
- [3] 高橋知也: 『セイバーメトリクスについての分析』, 南山大学数理情報学部数理科学科卒業論文, 2009.
- [4] 日本野球機構オフィシャルサイト
<http://www.npb.or.jp/>
- [5] 野球統計サイト Fangraphs
<http://www.fangraphs.com>