

# 日本のプロ野球のセイバーメ リクス指標による評価の考察

香取研究室：野田達也

## セイバーメトリクス

野球ライターで野球史研究家・野球統計の専門家でもあるビル・ジェームスによって1970年代に提唱されたもので、メジャーリーグのデータをもとにして統計学的に作られたもの。アメリカでは一般的に使われている。

$$\text{OPS} = (\text{安打} + \text{四球} + \text{死球}) \div (\text{打数} + \text{四球} + \text{死球} + \text{犠飛}) \\ + (1 \times \text{単打} + 2 \times \text{二塁打} + 3 \times \text{三塁打} + 4 \times \text{本塁打}) \div \text{打数}$$

打者が得点増に有効な打撃をしているかどうかを表す指標。数値が高いほどチームの得点増に貢献している打者だと判断できる。

- $$\begin{aligned} \text{XR} = & 0.50 \times \text{単打} + 0.72 \times \text{二塁打} + 1.04 \times \text{三塁打} + 1.44 \times \text{本塁打} \\ & + 0.34 \times (\text{四球} - \text{故意四球} + \text{死球}) + 0.25 \times \text{故意四球} \\ & + 0.18 \times \text{盗塁} - 0.32 \times \text{盗塁死} - 0.09 \times (\text{打数} - \text{安打} - \text{三振}) \\ & - 0.098 \times \text{三振} - 0.37 \times \text{併殺打} + 0.37 \times \text{犠飛} + 0.04 \times \text{犠打} \end{aligned}$$

XRは打者が生み出した得点を表す指標で、打撃を得点化したもの。

例えばXRが50ならその打者が得点50を生み出したということであり、チームとしてのXRは実総得点と近くなる。

- $$\text{BABIP} = (\text{安打} - \text{本塁打}) \div (\text{打数} - \text{三振} - \text{本塁打} + \text{犠飛})$$

本塁打を除き、グラウンド上に飛んだ打球が安打になった割合を示す指標。

•  $WHIP = (\text{被安打} + \text{与四球}) \div \text{投球回}$

イニングあたりにどれだけ走者を許したかを表す指標。

•  $LOB\% = (\text{安打} + \text{四死球} - \text{得点}) \div (\text{安打} + \text{四死球} - 1.4 * \text{本塁打})$

出塁した走者に得点を許さなかった割合を示す指標。

•  $FIP = \{13 \times \text{被本塁打} + 3(\text{与四球} - \text{故意四球} + \text{与死球})$

$- 2 \times \text{奪三振}\} \div \text{投球回} + 3.12$

本塁打、四死球、三振といった投手のみの責任によって、優劣を比較する指標

この他にも様々な指標があるが今回はセイバーメトリクス指標において代表的なこれらだけを扱うことにする。

# 目的

- 近年アメリカで注目を浴びているメジャーリーグの統計データをもとにして作られたセイバーメトリクス指標による評価分析が、日本のプロ野球を評価分析する際にも有効かどうかを調べる。

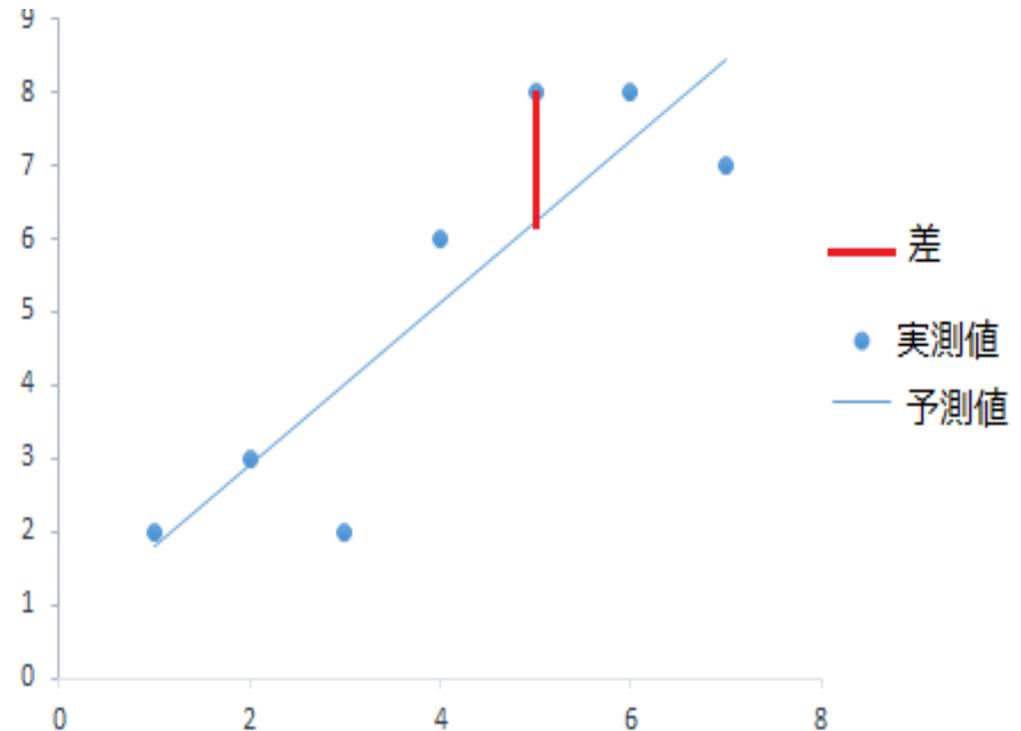
# 単回帰分析

- 重回帰分析とは、1つの従属(目的)変数を、1つの独立(説明)変数から予測・説明したいときに用いる統計手法

予測値  $\hat{y}_i = \beta_0 + \beta_1 x_i$

実測値  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

- $A^2 = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$



$A^2$ が最小になるようにとし値を決める。

$$\frac{\partial A^2}{\partial \beta_0} = 0 \quad \frac{\partial A^2}{\partial \beta_1} = 0 \quad \text{とし連立すると}$$

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

となる。

# 決定係数と相関係数

$$\bullet R^2 = 1 - \frac{\text{残差平方和}}{\text{偏差平方和}}$$

残差平方和  $A^2$

偏差平方和  $\sum_{i=1}^n (y_i - \bar{y}_i)$

分析の精度が良ければ残差平方和は0に近づき  
 $R^2$ は1に近づく。今回はR(相関係数)を使う。

また (相関係数)<sup>2</sup> = 決定係数

# 回帰分析結果

## 比較の方法

上記の式から関係のある項目で勝率と比較する。

## 例) OPSの場合

$$\text{OPS} = \frac{(\text{安打} + \text{四球} + \text{死球})}{(\text{打数} + \text{四球} + \text{死球} + \text{犠飛})} + \frac{(1 \times \text{単打} + 2 \times \text{二塁打} + 3 \times \text{三塁打} + 4 \times \text{本塁打})}{\text{打数}}$$

OPSと安打、単打、二塁打、三塁打、本塁打、犠飛、四球、死球、打数、を比較する

また打率 = 安打 ÷ 打数 なので打率とも比較する

- OPS(打者指標)

	安打	单打	二壘打	三壘打	本壘打	打率	OPS
勝率	0.318	0.140	0.287	0.300	0.227	0.382	0.400

- BABIP(打者指標)

	安打	单打	二壘打	三壘打	犠飛	BABIP
勝率	0.318	0.140	0.287	0.300	0.249	0.341

- XR(打者指標)

	得点	打点	XR
勝率	0.42	0.431	0.432

## ・ 投手指標の比較(相関係数)

### 日本のプロ野球の指標

	防御率	被安打	与四球	奪三振	被本塁打	与死球	失点	与故意四球
勝率	-0.620	-0.546	-0.344	0.352	-0.352	-0.098	-0.617	-0.192

### セイバーメトリクス指標

	WHIP	LOB%	FIP
勝率	-0.634	0.677	0.529

# 結果の考察

以上の結果より打者指標では日本のプロ野球で使われている指標よりもセイバーメトリクス指標の方が相関係数が高く、勝率を予測する際に有効であると言える。またその中でも特にXRの相関が高く、打者指標の中ではXRから勝率を予測することが望ましいと言える。

投手指標ではFIPの相関係数は防御率や失点よりも低い結果となったがWHIPやLOB%は相関が高くセイバーメトリクス指標の方が勝率を予測する際に有効であると言える。

これらよりメジャーリーグの統計データをもとにして作られたセイバーメトリクス指標は日本のプロ野球でも有効に使えるということがわかった。

# 日本版XRの公式の作成

- セイバーメトリクス指標について調べていくとXR指標は得点を予測するために、重回帰分析によりそれぞれの係数が求められているということがわかった。そこで独自に日本版XRを作成してみる。
- $$\begin{aligned} \text{XR} = & 0.50 \times \text{単打} + 0.72 \times \text{二塁打} + 1.04 \times \text{三塁打} + 1.44 \times \text{本塁打} \\ & + 0.34 \times (\text{四球} - \text{故意四球} + \text{死球}) + 0.25 \times \text{故意四球} \\ & + 0.18 \times \text{盗塁} - 0.32 \times \text{盗塁死} - 0.09 \times (\text{打数} - \text{安打} - \text{三振}) \\ & - 0.098 \times \text{三振} - 0.37 \times \text{併殺打} + 0.37 \times \text{犠飛} + 0.04 \times \text{犠打} \end{aligned}$$

# 重回帰分析

重回帰分析とは、1つの従属(目的)変数を、複数の独立(説明)変数から予測・説明したいときに用いる統計手法

$$\text{予測値 } \hat{y}_i = \beta_0 + \sum_{l=1}^p \beta_l x_{il}$$

$$\text{実測値 } y_i = \beta_0 + \sum_{l=1}^p \beta_l x_{il} + \varepsilon_i$$

単回帰分析と同様に

$A^2 = \sum_{i=1}^n \varepsilon_i^2$  とおきこれを最小にするような  $\beta_0, \beta_1, \dots, \beta_p$  を求める。

- データをもとに目的変数を得点とし、  
(目的変数とは上の式でいう $y$ のこと)  
変数を単打、二塁打、三塁打、本塁打、  
(四球－故意四球＋死球)、故意四球、盗塁、盗塁死、  
(打数－安打－三振)、三振、併殺打、犠飛、犠打として重回帰分析する。  
(上の式でいう $x_{il}$ のこと)  
これにより日本版の独自のXRを求める。

重回帰分析の結果を次の表に示す。

# 分かりやすいようにそれぞれの係数を表にする

	XR	日本版独自XR
単打	0.50	0.47
二塁打	0.72	0.79
三塁打	1.04	1.32
本塁打	1.44	1.41
四球－故意四球＋死球	0.34	0.36
故意四球	0.25	0.09
盗塁	0.18	0.22
盗塁死	-0.32	-0.35
打数－安打－三振	-0.09	-0.10
三振	-0.98	-0.11
併殺打	-0.37	-0.28
犠飛	0.37	0.75
犠打	0.04	-0.05

# 結果の考察

故意四球、盗塁、盗塁死、併殺打、犠飛、犠打はT値の絶対値が低く（有意水準1%）信憑性にかけるため考察からは除外する。

三振の係数が差が大きく0.87違い日本よりもメジャーリーグは三振が大きく影響しているということがわかる。

また一年毎のチームの得点の平均は566.8

一年毎のチームのXRの平均は599.4

一年毎のチームの独自の日本版XRの平均は566.8であり

セイバーメトリクス指標のXRは約30点過大に評価しているということがわかる。

## まとめ

これらより勝率を予測する際にセイバーメトリクス指標は打率や防御率といった日本で使われてきた指標よりも有効であったが、XRのようにメジャーリーグの統計データをもとにして作られたものに当てはめた場合過大に評価し過ぎるといったようなずれが生じてくることがわかった。

日本のプロ野球をより正確に評価するためには日本版のセイバーメトリクス指標を作る必要があると思われる。

## <参考文献>

「プロ野球本当の実力がわかる本ーセイバーメトリクスで見るプロ野球」

京都純典 2009年 日刊スポーツ出版社

「Excelで学ぶ多変量解析入門」 菅民郎 株式会社オーム社

日本野球機構オフィシャルサイト <http://www.npb.or.jp/>

日本野球機構オフィシャルサイトより2005年から2012年のチームのデータを使用