プロ野球データの統計的分析

2006MI027 日比崇允
指導教員：田中豊

1　はじめに

メジャーリーグでは, 数十年前からセイバーメトリクスという統計手法を用いて球団の構成や選手の評価を行っている. しかし日本プロ野球では, この統計手法に対して認識は決して大きくなく,「打率」,「安打」,「打点」などの従来の指標を参考にしているのが実情である. 本研究では,2008 年度の規定打席に達している両リーグの打者に注目し, 従来の指標とセイバーメトリクスを用いてどういう項目がどのように年俸に反映されているかについて分析を行った.

2　データについて

従来の指標とセイバーメトリクスの両リーグの 2008 年度の規定打席達成者の成績のデータと 2008 年度,2009 年度の年俸を用い, また,2008 年度の規定打席達成者から,2009 年度も規定打席に達している選手の,2009 年度の成績と 2010 年度の年俸のデータも用いた.

3　セイバーメトリクス

従来の指標では試合状況などに左右されることが多いのだが, それでは選手の能力を正確にはかることができない. この問題を解決してくれるのがセイバーメトリクスと言われている.

4　分析方法

分析方法としては, 主として重回帰分析を用いた. 重回帰分析とは, ある変数 y(従属変数) と, それに影響を及ぼすと考えられるほかの変数 $x_1, x_2, \cdots, x_p$(説明変数) に関するデータに基づいて,

$$y_i = a_0 + a_1 x_{i1} + \cdots + a_p x_{ip} + e_i (i = 1, 2, 3, \cdots, n)$$

という線形回帰モデルを仮定して, 回帰係数 $a_0, \cdots, a_p$ を推定する. 推定された回帰係数を用いて, 説明変数から x 従属変数 y を予測する方法である. 本研究では, 説明変数に対して変数選択法 (ステップワイズ法) を用いた.

5　年俸の傾向

図 1 は 2008 年度の年俸と 2009 年度の年俸の散布図である. この図を見てみると,2008 年度の年俸より 2009 年度の年俸のほうが高くなっているように見える. これらの年俸の変化に差があるかどうかを判断するために, 対応のある t 検定を行ってみたところ,p 値が $4.046 * 10^{-8}$ となり, これらの年俸の変化が有意であることがわかった. また, その差の平均金額が約 2500 万円であった. しかし多額の年俸をもらってる選手とそうでない選手がいるため, この金額が正しいとはいえない. そこで, 対応のある t 検定を行い, 割合で示したところ,p 値が $3.056 * 10^{-7}$ となり, その差の割合の平均は約 0.5 となった. これらのことから, 規定打席達成者の来年度の年俸は, その年の年俸の約 1.5 倍となることがわかった.

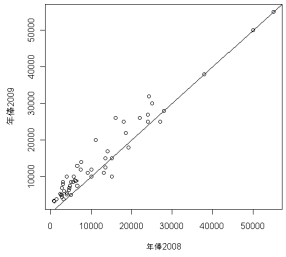

図 1　年俸の散布図

6　年俸予測

従来の指標とセイバーメトリクスのデータからそれぞれ 2009 年度の年俸を予測するモデルを作り,2 つの指標の違いを明確にするため,2009 年度の年俸を目的変数, その他の変数を説明変数として重回帰分析を行った. そして,2008 年度の規定打席達成者から,2009 年度も規定打席に達している選手の,2009 年度の成績と 2010 年度の年俸のデータを用いて, 年俸の予測モデルから,2009 年度の選手の成績を使って,2010 年度の年俸の予測値を出し, 実際の 2010 年度の年俸にどの程度当てはまっているのかを調べた.

6.1　重回帰分析の結果

説明変数としては, 従来の指標とセイバーメトリクスの指標と前年度の年俸, 年齢を用いた. 重回帰分析の結果をそれぞれ表 1, 表 2 に示す. ただし有意水準 5 %以下で効いている変数と, 前年度の年俸や年齢などは省略した.

表 1　従来の指標の重回帰分析の結果

	回帰係数	p 値	
二塁打	$-3.831 * 10^2$	0.000462	***
三塁打	$-4.550 * 10^2$	0.009978	**
本塁打	$-6.502 * 10^2$	0.006155	**
犠打	$-2.316 * 10^3$	0.009831	**
犠飛	$-2.836 * 10^3$	0.002433	**
四球	$-2.394 * 10^3$	0.007886	**
死球	$-2.424 * 10^3$	0.007187	**
長打率	$1.470 * 10^5$	0.000506	***
R^2	0.9831		

表 1 の結果では, 私はホームランをガンガン打てる打者が多額の年俸をもらっていると予想していたが, 実際は, 本塁打よりも二塁打や長打率などのヒットを量産できる

表 2 セイバーメトリクスの重回帰分析の結果

	回帰係数	p値	
RCAA	$-2.301 * 10^4$	0.000869	***
RCWIN	$1.886 * 10^5$	0.005602	**
XR27	$8.154 * 10^4$	0.005881	**
XR. 打席	$-2.259 * 10^6$	0.006276	**
XR.	$2.507 * 10^4$	0.000829	***
XRWIN	$-2.050 * 10^5$	0.003591	**
SecA	$6.862 * 10^4$	0.005718	**
IsoD	$-1.320 * 10^5$	0.002518	**
R^2	0.979		

タイプの指標と年俸が関係が深いといえる. また, 犠打や犠飛など味方を援護するタイプの指標や, 死四球などの打たずして出塁できるタイプの指標も年俸と関係が深いことがわかった. 表 2 の結果から,OPS や IsoP などの指標が年俸と深い関係にあると予想していたが, 実際は, 年俸は, 全体的に RCAA や XR. といった選手自身の得点能力を表す指標と関係が深いといえる. 従来の指標と違い,OPS や IsoP といった打者のヒットなどを重視した指標は, 他の指標と比べると関係が深いとはいえなかった.

表 3 表 2 の分析の残差を目的変数とした重回帰分析

	回帰係数	p値	
二塁打	-189.44	0.00145	**
本塁打	-304.51	0.00406	**
長打率	59690.22	0.00231	**
出塁率	-46537.05	0.00978	**
R^2	0.1214		

表 3 はセイバーメトリクスの残差を従来の指標でどの程度説明できるのかを表している.「二塁打」,「本塁打」,「長打率」,「出塁率」が有意水準 1 %で効いているという結果が出た. この結果から, セイバーメトリクスの欠点は純粋に長打や出塁するということを評価することができないのではないかと考えられる.p 値が 0.08096 であり, 自由度調整済み決定係数が 0.1214 という数値が出たため, 多少の影響力はあるといえる. また, 表 3 とは逆の分析も行ったが,p 値が非常に大きくなり, どの変数も効いているとはいえなかった.

6.2 2010 年度の年俸予測

図 2, 図 3 の散布図を見ると,2 つとも線形的な右上がりになっているように見えるので, 予測モデルの当てはまりは良いといえる. これらの散布図は,45 度の線より上にある番号の選手は, 予測された年俸よりも実際の年俸のほうが多くもらっていることを表し,45 度の線より下にある番号の選手は, その逆を表す. ここで,2 つの散布図の番号の位置に特徴のある 2 番の選手 (稲葉篤紀選手) と,10 番の選手 (栗原健太選手),15 番の選手 (田中浩康選手) に注目してみる. 稲葉選手と田中選手は, あまり大きな差があるとは言えないが, 従来の指標での成績では, 予測値よりも

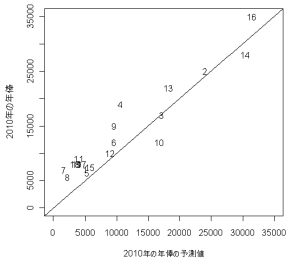

図 2 従来の指標の予測値と実際の年俸の散布図

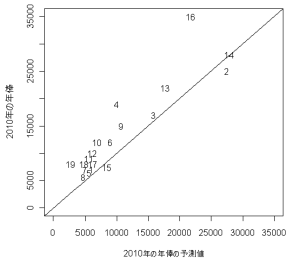

図 3 セイバーメトリクスの予測値と実際の年俸の散布図

少しだけ年俸をもらいすぎている, ということがわかり, セイバーメトリクスの成績では, もっと年俸をもらってもよいのではないか, という結果が出た. それとは逆に, 栗原選手は, 従来の指標での成績では, もっと年俸をもらってもよいのではないか, ということがわかり, セイバーメトリクスの成績では, 予測値よりも少しだけ年俸をもらいすぎている, ということがわかり, その差額が約 1 億円にもなった.

7 おわりに

今回の研究の結果から, プロ野球選手は規定打席に達することによって, 年俸がどのような変化をしているのかを知ることができた. そして, 従来の指標とセイバーメトリクスを使い, 年俸を予測するモデルを作ることによって,2 つの指標は年俸とどのような関係があるのかがわかった. また, 作ったモデルを利用し,2009 年度の成績から 2010 年度の年俸の予測値を出してみたところ, 比較的当てはまっていたので, うまく予測することができたと思う. しかし, 今回の研究のデータが, 規定打席達成者の成績と年俸のみを使用していたため, データ数が非常に少なかったので, データ数をもっと増やせば, より鮮明に結果が出て, 新しい発見もできたのではないかと思う.

参考文献

[1] 北島康博：『日本人メジャーリーガーの統計的分析』. 南山大学数理情報学部数理科学科卒業論文 2008.

[2] こちら, プロ野球人事部 http://home.a07.itscom.net/kazoo/pro/pro.htm

[3] 球団別インデックス http://www.npb.or.jp/teams/

[4] プロ野球データリーグ http://npbdl.web.fc2.com/