

多変量解析によるプロ野球選手の次期年俸推定（投手編）

本間 宏利* 石井 裕子**

A Forecast for Annual Salary of Professional Baseball Players (the Pitching Staff) with Multivariate Analysis

Hirotoishi HONMA* Yuko ISHII**

Abstract- Annual salary of professional baseball players is usually decided by negotiating between each player and a corporation owning a professional baseball team. In this study, we analyzed primary factors that determining annual salary and moreover constructed a formula for forecasting annual salary of each player with multivariate analysis.

Keywords: Multivariate Analysis; Correlation Coefficient; Multiple Linear Regression Analysis

1 はじめに

プロ野球選手の次期年俸は、年末に球団側と各選手が交渉することにより決定される。この時、そのシーズンの成績が大きな材料として翌年度の年俸額の決定に吟味されているが、このようにして決定される年俸は各選手によって大きな差があり、主力選手と控え選手の間には何倍もの差が生じている。ところが、選手によっては、貢献度に対して年俸額が少なすぎると評価される選手や、逆に多すぎるのではと非難されるケースが多く見られる。プロ野球選手は球団側に何の成績をどのような基準を持って評価をされて年俸額を決定されているのだろうか。このような背景を基に本研究では、読売巨人軍の選手（野手）を対象として、過去数年間の各選手の成績および年俸額から、年俸額を決定している要因の分析と、それらの具体的な評価を数式として一般化する年俸予想式の導出を多変量解析の技法を用いて試みた。昨年度の研究では読売巨人軍の野手を対象にした年俸予想式の導出 [4] を行ったが、今年度は投手を対象にした年俸予想式の導出とその評価を行う。

2 プロ野球選手の年俸の特性

プロ野球選手の翌年の年俸は前年度のシーズンの成績や活躍によって大きく影響を受けている。しかし、過去数年の選手個人の成績とその年俸額を分析した結果、必ずしも単純に前年度の成績に比例した年俸額が割り当てられているとはいえない。例をあげると、仮に本塁打数が30本の選手が同じく10本の選手の3倍の年俸をもらっているかというそうではなく、実際にはそれ以上の評価をされている。また、年俸が数百万円の新人選手が本塁打を20本打ったとしても、翌年の年俸がすぐに1億円以上になることは少なく、これに対して、毎年20本ずつ本塁打を5年間打ち続けている選手は数億円の年俸をもらっている。このように、非常に優れた成績を残したとしても、すぐに高年俸とはならず、逆に数年間に

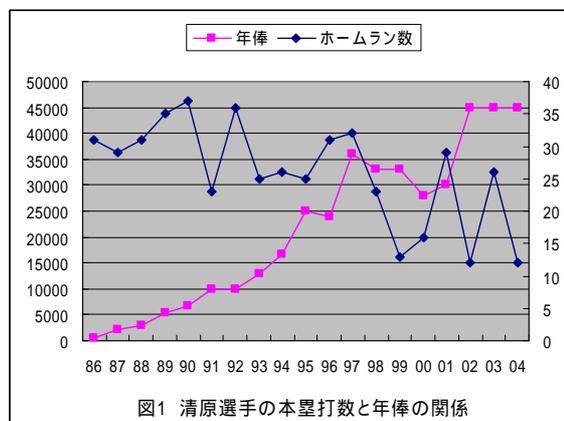


図1 清原選手の本塁打数と年俸の関係

* 釧路高専情報工学科 ** 株式会社ポータス

渡ってある程度以上の安定した実績を残した選手には成績以上の評価がされているといった傾向が強く見られる。具体的な例として図1に読売巨人軍の清原選手の本塁打数と年俸の関係を示す。清原選手といえりリーグを代表するホームランバッターであり、彼のプロ野球選手としての価値は本塁打数で評価されているといってもよいだろう。清原選手の本塁打数と年俸の関係を見てみると、入団当初は30本以上の本塁打数を成績として残しており、年俸額も徐々に高くなり、5年後には1億円を突破している。ところが、成績が少しずつ落ち始め、本塁打数がやっと20本台を維持しているような状況でも年俸は上昇し続け、現在では4億円以上の年俸をもらっている。もちろん、清原選手のプロ野球選手としての価値は本塁打数のみで決定される訳ではなく、特に清原選手の場合はフリーエージェント (FA) による移籍という経緯が年俸を高額にさせている事実もある。

この例からもわかるように、プロ野球選手は新人時代は年俸が低く、彼らの活躍によって徐々に年俸額は高くなる。そして、ある一定基準以上の安定した成績を数年に渡って残すようになると、年俸額が高騰するという傾向があることがわかる。

以下にプロ野球選手の成績と年俸の関係に見られる特徴を整理する。

1. 成績の数字と年俸額は単純に比例しない。
2. 1年の優秀な成績よりも、数年に渡る安定した成績が評価されやすい。
3. 前年度と同じ成績の場合、年俸が増えるか減るかはその選手によって異なる。
4. 優勝した翌年は年俸アップ率が大きい。
5. 個人タイトルを獲得した選手は年俸が高くなる。
6. 人気のある選手は年俸が高い。
7. 守備の評価は数字に表れにくい。
8. 選手によって評価される要因が異なる。
9. 年俸がある程度高い選手は成績が優れなかった年でも、翌年に大幅に年俸が下がることは少ない。

以上の関係1, 2, 3より、我々はプロ野球選手の次期年俸は前年度の年俸に、そのシーズンの成績によって算出されるある係数を乗じることによって決定されていると予測する。本研究ではこの係数のことを**年俸係数**と呼ぶことにする。すなわち、

$$\text{次期年俸額} = \text{年俸係数} \times \text{前年度の年俸} \text{ となる。}$$

3 年俸係数式 (投手) の導出

年俸係数の導出にあたって、それを決定する成績要因を分析する。選手 (投手) を評価する客観的な要因として、出場試合数、完投数、勝数、負数、セーブ数、勝率、投球回数、奪三振数、防御率、ボーク数、被安打数、被本塁打数、与四死球数等、様々なデータが存在する。これらの全ての要因を年俸係数の導出に取り入れることは、ある意味理想的ではあるが、冗長な要因を多く含むことや、数式が複雑になることから現実的ではない。そこで、はじめにこれらの各成績要因の全組合せに対して**相関係数**を導出し、年俸額と相関の強いと思われる要因を数点に絞る。さらに、それらの相関係数を基に、年俸額に対して冗長と思われる要因を削除、あるいは包括する。

具体的には勝数と年俸には高い相関関係が見られるが、多くの勝ち星を挙げている投手はそれだけ出場機会も多く、必然的に負数や被本塁打数、被安打数、与四死球数等の負の成績も多くなる。よって、負数や被安打数も年俸と相関が非常に高い要因となるが、これは、よく負ける投手程年俸が高くなるという意味ではない。このような分析作業および包括作業により、相関の強い要因の重複を排除した。この結果、**勝数(0.48)**、**勝率(0.24)**、**投球回数(0.48)**の3つの要素の数値が年俸額に大きな影響を与える要因になっていることがわかった。

しかしながら、この結果を吟味してみるとこれらの要素は全て先発投手を対象とした偏った評価基準になっていることに気づく。リリーフ投手や中継ぎ投手には上記のような要素の成績は貢献度として反映されにくい。これは、読売巨人軍は元来先発投手が豊富であり、反対に中継ぎ投手やリリーフ投手層が非常に薄く、あまり活躍している選手がいないという特色が映し出されている。より一般的には投手の評価として防御率やセーブ数を考慮すべきであり、本研究では実験的に上記の3つの要素の他に実験的に**防御率**と**セーブ数**を評価要因として加えることにした。

表2 要因間の相関係数 (投手)

	試	完	勝	S	率	回	三	防	年
試合	1.00								
完投	0.05	1.00							
勝	0.25	0.79	1.00						
S	0.40	-0.11	-0.02	1.00					
勝率	0.37	0.28	0.53	0.11	1.00				
投回	0.29	0.80	0.93	-0.03	0.44	1.00			
三振	0.33	0.72	0.90	0.03	0.44	0.97	1.00		
防率	-0.32	-0.23	-0.34	-0.12	-0.49	-0.34	-0.36	1.00	
年俸	0.15	0.18	0.48	0.11	0.24	0.46	0.49	-0.15	1.00

続いて、**勝数**、**セーブ数**、**勝率**、**投球回数**、**防御率**の成績から、年俵係数を求める関係式の導出を行う。最初に留意すべき点として、これらの5つの成績が示す数値を絶対的な評価として扱うと多くの矛盾が生じる場合があることを示す。例えば、ある選手がその年度のシーズン10勝したと仮定する。ここで、この選手が前年度は5勝程度しか勝てなかった新人選手であれば、その選手への評価は非常に高く、翌年の年俵も大きくアップすることになる。反対に、この選手が毎年15勝以上も勝利をあげる選手であったならば、当然高額な年俵をもらっていることから、その年俵額と期待に対して非常に不満な成績といえる。このように、同じ数値でもその選手（の年俵）によっては、評価が高くもなり、低くもなる。

よって、各選手の成績を絶対的な指標として評価するのではなく、前年度の成績と比較して、どの程度活躍をしたのか、または、活躍できなかつたかを相対的な指標として評価する必要がある。したがって、本研究では各選手の今年度の成績と前年度の成績の差を変数として年俵係数を導出する年俵係数式の導出を行う。ただし、この手法ではある選手が前年度と今年度で全く同じ成績を残したと仮定すると、その差がゼロであるため低い評価を受ける可能性がある。このことは大きな矛盾を生じる可能性がある。例えば、ある選手が2年連続で20勝をあげたとすると、同じ成績にもかかわらず、この選手の年俵は大きく増加するのが現状である。もし、この選手が2年目に15勝しかしなかったとしても年俵は現状維持以上になることが多い。よって、年俵係数を求めるにあたり、選手が昨年度と全く同じ成績を残したと仮定した時の年俵増加率をベース値として考慮する必要がある。

以上のことから、我々は選手の年俵係数式を次のように構築した。 x_1 を今年度の勝数と昨年度の勝数の差。同様に、 x_2, x_3, x_4, x_5 をそれぞれ、今年度と昨年度のセーブ数、勝率、投球回数、防御率との差とし、これらの変数にそれぞれ、 a_1, a_2, a_3, a_4, a_5 を重み係数として、積和をとる。この線形式に前年度と今年度の成績が全く同じだったと仮定した場合の年俵アップ率（ベースアップ値） b を加えることで年俵係数を求める。年俵係数式を下記に示す。

$$\text{年俵係数(投手)} = \sum_{i=1}^5 a_i x_i + b$$

x_1 : 勝数, x_2 : セーブ数, x_3 : 勝率, x_4 : 投球回数,
 x_5 : 防御率, b : ベースアップ値

この年俵係数式は、変数 x_i を**独立変数**、係数 a_i を**回帰係数**、ベース値 b を**定数項**とみなした場合に多変量解析手法の一つである**重回帰分析**を適用して最適な回帰係数および定数項を導出することが可能である。

4 年俵係数式の改善

プロ野球選手の年俵は数値として現れる成績のみによって決定されている訳ではない。例えば、守備力、走力、勝負強さ、人気、リーダーシップ性、スター性、個人タイトルの獲得等も年俵額に大きく影響する要素であると思われる。昨年の研究（野手編）でも数値に表れないこれらの要因を年俵係数式に組み入れることによって、より正確な予想年俵額の導出を実現できた。今回の投手編にもこれらの要素を組み入れる。

4.1 走力

走力が数値として現れやすいのは盗塁数であろう。また、塁上に走力のある選手がいた場合、1本のヒットで得点に結びつく可能性も高くなる。しかし、読売巨人軍にはここ数年、野手にも走力を評価される選手は少なく、ランナーとして塁上に立つ機会の少ない投手にはほとんど年俵として評価される数値ではないため除外する。

4.2 守備力

野手にとって守備力の評価は非常に重要である。攻撃面で活躍でなくても守備力で高い年俵を稼ぐ選手も少なくない。しかし、投手の場合は守備力といえばフィールディング能力に相当すると思われるが、その守備機会が非常に少ないことや、元々投手力=守備力という考えがあるため、この要因を年俵予想から除外する。

4.3 人気度

年俵に影響を与える大きな要素としてその選手の人気がある。実力があるので人気があると見なすのが正論ではあるが、現実的には人気ばかりが先行して成績が伴わない選手が数多く存在するのは事実である。しかし、球団側の観点から考えると、人気のある選手が試合に出場することにより、集客力が増え、視聴率が上がり、メディアが取り上げてくれることで大きな宣伝効果となるため、そのような選手に高い年俵を与えても経営は成立する。よって、人気という要因を独立変数として年俵係数式に取り入れることにする。選手の人気度を可能な限り客観的に数値化するため、本校の学生にアンケートにより、人気があると思われる選手に投票をしてもらい、そのデータを正規化して人気度数を得た。各選手の人気

度数を表3に示す。

表3 各選手の人気度数

選手名	評価	選手名	評価
河原	1.02	工藤	2.65
高橋	2.74	柏田	0.11
桑田	4.21	酒井	0.05
上原	5.00	真田	0.60
河本	0.32	鴨志田	0.08
岡島	0.92	木村	0.06

4.4 個人タイトル

選手に与えられる名誉な記録として個人タイトルがある。選手と球団間での契約の際には、何らかの個人タイトルを獲った選手には、通常の年俸の他にタイトル料として更にボーナスが付帯される契約を結ぶ場合が多い。この個人タイトルも様々な種類があるが主なものとして投手三冠と呼ばれる最多勝利、最優秀防御率、最優秀勝率、最多セーブを獲得した選手には特別に年俸アップがあると予想され、多くのメディアでは1つのタイトルに付き約1,000万円程度とみなされている。よって、これまでに入力データとしていた各選手の年俸データに対して、個人タイトルを獲得した選手の翌年の年俸から獲得タイトル1つに付き、1,000万円を減じた数値をより信頼性の高いデータとして扱う。また、個人タイトルを獲得した選手には、年俸予想式で算出された年俸にタイトル数だけ1,000万円を加算して予想年俸額とする仕様に変更する。

表4 個人タイトルボーナス

タイトル名	ボーナス額
最多勝利	1000万円
最優秀防御率	
最優秀勝率	
最多セーブ	

4.5 優勝ボーナス率

選手の毎年の年俸上昇率を調べると、リーグ優勝した翌年の年俸のアップ率が非常に高いことがわかる。優勝するという事は、各選手が高い水準の成績を残したから年俸アップ率が高いのは自然ではあるが、この年俸予想式で算出された年俸額より非常に高い年俸を各選手が受けている。この理由として、優勝を争うシーズンは観客動員数も多く、シーズン終盤まで視聴率も高くなり、また、日本シリーズ等のイベントなどで、より多くの収入を球団側は得る。よって、優勝という事実が成績以上の年俸アップを実現させていると

推測できる。

従って、我々は優勝した翌年の年俸は、仮に優勝しないで同じ成績を残した場合と比較して、どの程度の年俸上昇率(優勝ボーナス率)があるのかを実験的に求める。まず、優勝ボーナス率を1%から50%までと想定して、過去の優勝した翌年の年俸データからここで想定した優勝ボーナス率に対する金額を排除した年俸額と、我々の年俸予想式から算定された年俸額との差が最小になる優勝ボーナス率を求めた。図5の横軸は想定した優勝ボーナス率を示し、縦軸はその優勝ボーナス率による年俸上昇額を排除した年俸額と、年俸予想式から算定される年俸額との差である。よって、この実験により、優勝したシーズンの翌年には通常年俸上昇の他に、約28%アップの優勝ボーナスが与えられていると予想される。

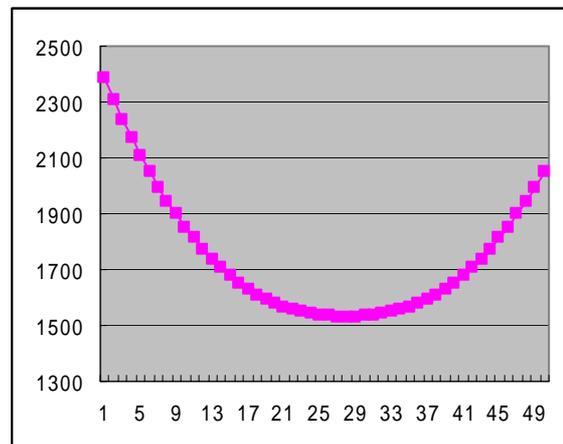


図5 優勝ボーナス率

5 年俸予想結果と評価

本研究で導出した年俸係数式とそれによって導出された各投手の予想年俸およびその精度の評価について記述する。年俸係数式の独立変数には読売巨人軍の全投手の数年に渡る成績を利用した。ここで、対象となる選手は一軍登録選手のみとし、イースタンリーグの成績は一切考慮はしない。また、オールスターや日本シリーズなどの成績も考慮せず、公式試合だけを対象にした。オールスターに出場し活躍した選手は人気度数が上がり、日本シリーズでの活躍は優勝ボーナス率に影響すると思われるため、ここでは省略しても問題ないと考えられる。また、トレードやフリーエージェントで入団した選手に対しては読売巨人軍に入団以降の成績だけを対象とした。改善した新たな年俸係数式を下記に示す。この年俸係数式には人気度を独立変数として導入した。

$$\text{年俵係数 (投手)} = \sum_{i=1}^6 a_i x_i + b$$

x_1 : 勝数, x_2 : セーブ数, x_3 : 勝率, x_4 : 投球回数,
 x_5 : 防御率, x_6 : 人気, b : ベースアップ値

この年俵係数式を基に, 重回帰分析によって導出された年俵係数式の回帰係数およびベース値を表 6 に示す。

表 6 回帰係数

勝数	a_1	0.02581
セーブ数	a_2	0.02057
勝率	a_3	0.05496
投球回数	a_4	0.00402
防御率	a_5	0.01580
人気	a_6	0.01739
ベースアップ値	b	1.22782

この回帰係数では勝率の重みが大きくなり, 次に勝数やセーブ数が大きい。逆に投球回数の重みが最小となっている。しかし, 勝数やセーブ数は元々数値の変動が小さく, 反対に投球回数は非常に大きな数値をとると予想されることから, 単純に勝率が最も年俵に影響を与えるという訳ではないことに注意する。また, 今年度の成績が昨年度の成績と全く同じだった場合を想定した年俵のベースアップ率は 1.227 倍となった。

この年俵係数式を基に, 各選手 (投手) の次期年俵を導出し, 実際の年俵との比較を行う。今回は 2003 年度の成績を基に 2004 年度の年俵を予想し, 実際の年俵との比較を行った (これは昨年行った野手の年俵予想と投手の年俵予想を比較するために 2004 年度の年俵予想に統一させた)。この結果を表 7 に示す。この予想額は個人タイトル獲得ボーナスと優勝ボーナスを考慮した年俵額である。

表 7 予想年俵と誤差 (万円)

選手名	予想年俵	誤差
河原	3420	-2580
高橋	6954	-646
桑田	19578	-1422
上原	29484	-516
河本	7945	1945
岡島	11341	1341
木村	4801	2101
工藤	31016	3016
柏田	4296	396
酒井	1381	181
真田	2444	-56
鴨志田	692	-198

予想年俵額の精度を求めめるため, 下記の 2 つの指標を用いる。ここで, 選手の人数を n , 各選手の予想年俵額を E_i , 実際の年俵を T_i とする。

$$\text{誤差額} = \sqrt{\frac{\sum (T_i - E_i)^2}{n}}$$

$$\text{誤差率} = \frac{1}{n} \sum \left| 1 - \frac{E_i}{T_i} \right|$$

この年俵係数式で導出した 2004 年度の投手の予想年俵額の精度を表 8 に示す。

表 8 2004 年度の予想年俵の精度

	2004 年度
誤差額	1546 万円
誤差率	20.35 %

6 考察

本研究では, 選手の年俵額と相関が高い勝数, セーブ数, 勝率, 投球回数および防御率をもとに, 前年度と今年度の成績の差から翌年の年俵の上昇率である年俵係数を求め, 次期年俵額を予想した。数値に表れない要因として人気度を考慮し, 更に, 個人タイトル獲得ボーナスや優勝ボーナスを考慮して, 次期年俵を予想し, その精度を調べた。

ここで, 昨年度に行った巨人軍野手の次期年俵予想との精度の比較を行いその結果を表 9 に示す。野手の係数の導出には, 要因として安打+四死球数, 本塁打数, 打点数, 犠打数, 人気度, 守備力を用いた。

表 9 野手と投手の予想年俵の精度

選手名	野手	投手
誤差額	2235 万円	1546 万円
誤差率	16.75 %	20.35 %

この結果では平均の誤差額では投手の予想が優れているが誤差率では野手の予想が優れている。これは野手の方が投手に比べて年俵額が非常に大きい選手が数多く存在することや, 野手の中でも年俵の格差が非常に大きいことが原因であると思われる。また, 誤差率で投手の予想が劣っている原因としては, 年俵と相関の低かった防御率とセーブ数を実験的に年俵係数式を導出するための係数に取り入れたことが原因であろう。巨人軍は

元来、優秀な先発ピッチャーが活躍し、中継ぎや抑え投手は活躍の場数が少なく、優秀な投手が育つ環境のチームとはいえないのが現実である。投手の予想年俵の精度を更に向上させるには、先発陣と中継ぎ抑え陣のそれぞれに対して年俵予想式を構築する必要があるだろう。

次に、表 10 に野手と投手のベースアップ率と優勝ボーナス率の算出結果を示す。ベースアップ率とはその選手が昨年と今年の成績が全ての要因で全く同じ結果だった場合に年俵がどのくらい上昇しているかを予想した数字である。また、優勝ボーナス率とはリーグ優勝した翌年度の選手の年俵は優勝できなかった年度に比べてアップ率がよいという事実から、そのアップ率を算出した数値である。

表 10 ベースアップ率とボーナス率

選手名	野手	投手
ベースアップ率	1.05 %	1.23 %
優勝ボーナス率	17 %	28 %

結果としては投手の方がベースアップ率も優勝ボーナス率も野手より非常に高く設定されていることがわかった。つまり、2年間同じ成績を残すということは、投手にとって非常に大きな仕事をしたとみなされることであり、言い換えると2年連続して安定した成績を収めることは投手にとって非常に難しいことを表している。また、優勝ボーナス率も野手と比較して格段に大きな値が算出された。このことは、優勝したということは投手陣の働きが非常に大きかったということであり、年間を通してよいチーム成績を残すことは投手の成績次第と考えることができる。事実、ここ数年巨人軍は至上最強打線を売り物にしながらも、投手陣が打ち崩されることによって優勝から遠ざかっている。

7 今後の課題

今後の課題として、年俵係数式を改善し、より精度の高い年俵予想を行うことが必要である。このためには以下のような成績以外の要因をどのように取り入れるかが課題の一つになる。

1. トレード・FAによる年俵の高騰。
2. 勝負強さ（勝利勝率、得点圏打率等）の評価。
3. リーダーシップ性の評価。
4. 公傷などによる試合欠場の評価。

また、年俵が高い選手と低い選手にグループ化し、それぞれに対して最適な年俵係数式を導出させることや、日本シリーズ、オールスターの活躍や、メディアアピール等の球団貢献度などの評価の組み入れも課題になる。

- [1] 大村 平，多変量解析入門，啓学出版，1987
- [2] 長谷川勝也，これならわかる多変量解析，技術評論社，2001
- [3] 読売巨人軍オフィシャルサイト，<http://www.giants.jp/>
- [4] 本間宏利，石井裕子，高山木綿，多変量解析によるプロ野球選手（読売巨人軍）の次期年俵推定，釧路高専紀要，2004