## 多変量解析によるプロ野球選手(読売巨人軍)の次期年俸推定

本間 宏利\* 石井 裕子\*\* 高山 木綿\*\*\*

# A Forecast for Annual Salary of Professional Baseball Players (YOMIURI GIANTS) with Multivariate Analysis

Hirotoshi HONMA Yuko ISHII Yu TAKAYAMA

**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契約される. この時, そのシーズンの成績が大きな材料として次期年俸額の 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のデータから, 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.

#### 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ことは容易に予想できる.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. 例を挙げれば、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とそうではなく、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ある。具体的な例として図1に元読売巨人軍の松井選手の本塁打数と年俸の関係を示す。松井選手は新人時代は成績の割に年俸が低いが、安定した成績を残すようになってから年俸が急騰していることがわかる。

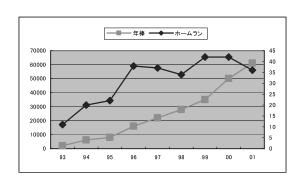


図 1 松井選手の本塁打数と年俸の関係

<sup>\*</sup> 釧路高専情報工学科 \*\*株式会社ポータス \*\*\*釧路高専第13期情報工学科

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記述する.

- 1. 成績の数字と年俸額は正比例しない.
- 2. 単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. 前年度と同じ成績でも選手により年俸は増減する.
- 4. 優勝した翌年は年俸アップ幅が大きい.
- 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からといって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、

次期年俸額 = 年俸係数 × 前年度の年俸 となる.

## 3. 年俸係数, 年俸予想式

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、選手の活躍を評価する客観的な数値尺 度として, 出場試合数, 安打数, 打率, 打点, ホーム ラン, 三振数, 盗塁数, 犠打数, 四死球数, 長打率, 得点圏打率等、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み、数式 が複雑になることから現実的ではない、よって、最初 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. 具体的には表 2 に示すように、活躍している選手は出 場試合数も多く、当然安打数も多くなり、打点や得点 も多くなるため、これらの要因間には強い正の相関が 見られる. また、これらの要素と三振数との間にも強 い正の相関が見られるが、これは活躍している選手ほ ど三振の機会も多くなるということを意味しており、 決して三振する打者ほど年俸が高いという訳ではない. 同様にホームラン数と打点数の間にも強い正の相関が 見られるが、ホームランの多い打者ほど打点も多くな るのは当然の結果といえる. また, 強打者ほど必然的 に四死球数も多くなるが、その数は実質的には安打数 に相当すると考えられる. このような分析作業により, 相関の強い要因の重複を排除し、四死球数と安打数の

包括化を行った.この結果, **安打+四死球数, 本塁打数, 打点数, 犠打数**の4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.

表2 要因間の相関係数 (野手)

	試	ŧΤ	得	安	本	点	盗	四	犠	Ξ	率	年
打数	0,98											
得点	0.89	0.79										
安打	0.98	0.97	0.88									
本塁	0.72	061	0.92	0.73								
打点	0.90	0.82	0.98	0.88	0.90							
盗塁	0.33	0.45	0.07	0.34	-009	0.20						
辎打	0,56	0.54	0.50	0.52	029	0.53	0.34					
四死	0.82	0.70	0.94	0.79	086	0.93	-002	0.37				
三振	092	0.91	0.86	0.93	0.69	0.81	0.18	0.49	0.71			
打革	0.89	083	0.92	0.96	085	0.90	0.13	0.48	086	0.85		
年数	0.63	0.48	0.86	0.64	0.88	0.80	-0.29	0.17	0.90	0.60	0.81	
年俸	0.53	0.36	0.78	051	0.74	0.71	-0.29	0.65	0.84	0.50	0.73	0.98

続いて、安打+四死球数、本塁打数、打点数、犠打数の成績から、年俸係数を求める関係式の導出を行う。最初に留意するべき点として、これらの4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。例えば、ある選手がその年度のシーズンに本塁打を20本打ったと仮定する。ここで、この選手が前年度は5本程度しか本塁打を打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、翌年の年俸も大きくアップすることになる。反対に、この選手が毎年40本以上も本塁打を打つ主力選手であったならば、当然高額な年俸をもらっていることから、その年俸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(の年俸)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(数値)を絶対的な指標として評 価するのではく、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、活躍できなかったかを相対的な 指標として評価する必要がある. したがって、本研究で はそれぞれの選手の今年度の成績と前年度の成績の差 を変数として年俸係数を導出する線形式(年俸係数式) の構築を行う. 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残す. 例えば, ある選手が2年連続 で 40 本本塁打を打ったとすると、同じ成績にもかかわ らず、この選手の年俸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30本しか本塁打を打たなか ったとしても年俸は最低限、現状維持に収まることが多 い. よって、年俸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俸増加率を ベース値として考慮する必要がある.

以上のことから,我々は選手の年俸係数式を次のように構築した。 $x_1$ を今年度の安打数+四死球数と昨年度の安打数+四死球数の差。同様に、 $x_2$ 、 $x_3$ 、 $x_4$ をそれぞれ,今年度と昨年度の本塁打数,打点数,儀打数との差とし,これらの変数にそれぞれ, $a_1$ , $a_2$ , $a_3$ , $a_4$ 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俸アップ率 (ベース値) b を加えることで年俸係数を求める. 年俸係数式を下記に示す.

年俸係数(野手) = 
$$\sum_{i=1}^{4} a_i x_i + b$$

 $x_1$ : 安打+四死球数,  $x_2$ : 本塁打,  $x_3$ : 打点 $x_4$ : 犠打, b: ベース値

この年俸係数式は,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選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野手の数年にわたる成績を利用した。ここで、対象となる選手は一軍選手のみとし、イースタンリーグ(2軍の試合)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。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。

重回帰分析によって導出された年俸係数式の回帰係 数およびベース値を下記に示す.

表 3 回帰係数

安打+四死球	$a_1$	0.00243004
本塁打	$a_2$	0.00208999
打点	$a_3$	-0.00174018
犠打	$a_4$	-0.00913622
ベース値	b	1.25003213

この年俸係数式を基に、各選手(野手)の次期年俸を導出し、実際の年俸との比較を行う。今回は2003年度の成績を基に2004年度の年俸を予想し、実際の年俸との比較を行った。予想年俸額の精度を求めるため、下記の2つの指標を用いる。ここで、選手の人数をn、各選手の予想年俸額をE, 実際の年俸をT<sub>i</sub>とする。

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

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

この年俸係数式で導出した 2004 年度の選手の予想年 俸額の精度を下記に示す.

表 4 2004 年度の予想年俸の精度(その1)

	2004 年度
誤差額	2787 万円
誤差率	20.09 %

この年俸係数式にはいくつかの問題点がある. 打点と 犠打の回帰係数に負の数値がついており、これは打点や 犠打数が増えるほど、年俸額が減少するという矛盾を含 んでいる. また、予想年俸の精度としても、誤差額、誤 差率ともに高精度といえる数値ではないことがわかる.

## 5. 更なる改善

これまでの年俸予想の問題点を考察し、より精度の高い予想年俸式の構築を目指す.プロ野球選手の年俸は数値として現れる打撃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俸額に大きく影響する要素であると思われる.よって、これら個々の要素について年俸予想式に反映させることが必要である.

#### 5.1 走力

まず、走力であるが単に足が速いだけでは意味がなく、 試合に出場し、盗塁を重ねることにより評価が高くなる。 今回の年俸係数式の要素の中に盗塁の要素が存在しな い理由は、読売巨人軍に盗塁の得意な選手がほとんどい ないということや、盗塁の得意な選手は若手選手に多い が、そのような選手は年俸が安く、年俸の高い選手はほ とんどがベテラン選手であることなどがあげられる。

#### 5.2 守備力

守備力は数値に表れにくい要素である。単純にエラー数で評価できるとも思えるが、守備機会数やポジションの違いによってその評価は非常に困難である。例えば、セカンドやショートは打球がよく飛んでくることが多く守備機会が増える上に、中継プレイにも加わるケースが多い。よって、必然的にエラーの数も多くなる。また、守備の上手な選手ほど守備範囲も広くなり、積極的なプレイも多くなることから通常ヒットと記録される場面でも記録上、エラーになることも多くなる。更には、内野手のエラーと外野手のエラーではチームに与える損失が大きく異なるという問題もある(一説では、外野手のエラーは内野のエラーの2倍に相当すると言われている)。その他にも、走力や肩の強さやキャッチングセンス等も守備力には大きく関係する要素であり、これらを

数値で表現することは困難といえる.よって守備力は 独立変数として年俸係数式に付加するが,その数値は 野球に詳しい人物に1から5の数値によってランク付 けを行ってもらう方法を用いることにした.表5に各 選手の守備力の数値を示す.

表 5 各選手の守備力の評価

選手名	評価	選手名	評価
元木	3	福井	2
川中	3	小田	2
後藤	3	阿部	4
清原	2	村田	3
二岡	4	清水	3
仁志	4	高橋	5
江藤	1	斉藤	3

#### 5.3 人気度

年俸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俸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俸係数式に取り入れることにした。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を行い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6に示す。

表 6 各選手の人気度数

選手名	評価	選手名	評価
元木	2.08	福井	0.12
川中	0.17	小田	0.03
後藤	0.18	阿部	1.13
清原	3.96	村田	0.06
二岡	0.36	清水	0.41
仁志	1.24	高橋	3.74
江藤	0.42	斉藤	0.03

## 5.4 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。選手と球団間での年俸契約時には、毎年、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三冠と呼ばれる打点王、首位打者、本塁打王を

獲得した選手には特別に何らかの形式で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した.

#### 5.4 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 するということは、各選手が高い水準の成績を残したか ら年俸上昇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. 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 させていると推測することは容易である.

よって、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30%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7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の年俸上昇の他に、約17%アップの優勝ボーナスが与えられていると予想される。

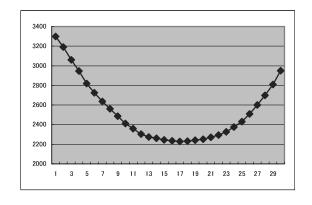


図7優勝ボーナス率

これらの改良によって新たに構築した年俸係数式を 下記に示す.この年俸係数式には新たに人気度と守備 力を独立変数として導入した.

年俸係数(野手) 
$$=\sum_{i=1}^6 a_i x_i + b$$

 $x_1$ : 安打+四死球数,  $x_2$ : 本塁打,  $x_3$ : 打点  $x_4$ : 犠打,  $x_5$ : 人気,  $x_6$ : 守備力, b: ベース値

この年俸係数式を基に,重回帰分析によって導出された年俸係数式の回帰係数およびベース値を下記に示す.

表 8 同帰係数 (改良)

20		(5/1/)
安打+四死球	$a_1$	0.004715288
本塁打	$a_2$	0.020878602
打点	$a_3$	-0.003255261
犠打	$a_4$	0.005179926
人気	$a_5$	0.014309751
守備力	$a_6$	0.026128486
ベース値	b	1.053216636

上記の年俸係数予想式から2004年度の各選手の予想年俸額を算出し、更に、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俸額を導出した。この予想年俸額と実際の年俸額との誤差を表9に示す。また、年俸予想の精度を推定する誤差額と誤差率を改良前の誤差額、誤差率と比較させて表10に示す。

表9 予想年俸と誤差(万円)

選手名	予想年俸	誤差
元木	10515	-1365
川中	3189	189
後藤	5986	-1114
清原	44344	-656
二岡	13494	-3506
仁志	20316	5016
江藤	22208	708
福井	2319	419
小田	1629	126
阿部	8399	899
村田	4145	245
清水	12532	-3968
高橋	33114	3114
斉藤	5000	-500
堀田	1596	116

表 10 2004 年度の予想年俸の精度(その2)

	2004 年度	2004 年度(改良)
誤差額	2787 万円	2235 万円
誤差率	20.09 %	16.75 %

## 6. 考察

本研究では、最初に選手の年俸額と相関が高い安打+四死球数、本塁打数、打点数、犠打数を基に、前年度との成績の差から翌年の年俸の上昇率である年俸係数を求め、翌年の年俸額を予想した.数値に表れない要因として人気度や守備力を加え、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、年俸を予想した.これらの要因を加えた結果、より誤差の少ない年俸予想に成功した.しかし、打点に対する回帰係数を見ると、重みが小さい値とはいえ負の数値になっている.このことは、打点を多くあげればあげるほど年俸が下がることを意味している.明らかに矛盾であるが、これは以下の原因が挙げられるだろう.

- 1. トレード・FA による年俸の高騰による誤差.
- 2. 勝負強さ (勝利打点, 得点圏打率等) の評価.
- 3. リーダーシップ性の考慮.
- 4. 公傷などによる試合欠場の評価.

本研究の結果をふまえて今後の課題として、年俸式を 改善すること、年俸が高い選手と低い選手にグループ化 し年俸を予想していくことが挙げられる。また、野手の 他に、投手の年俸予想を行うことも課題である。

- [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/