

卒業論文

2019 年度プロ野球 統計的分析

大畑滉貴

慶應義塾大学経済学部

要旨

近年、送りバントの有効性が疑問視されたり、2番最強説が唱えられるようになったりと、従来のセオリーが覆されつつある。得点及び勝敗に強く影響与えているものを分析していくと安打や失策等が影響を与えていたのはもちろんのことだが、打撃結果に関わらず「2番打者」というだけでも影響を与えていることがわかった。

目次

1. はじめに	P1
2. 先行研究	P2
3. 分析	P2
4. 考察・まとめ	P11
6. 参考文献	P11

1. はじめに

平成最後や令和初など様々な冠がついた 2019 年度のプロ野球はパリーグを埼玉西武ライオンズ、セリーグを読売巨人軍が制し、プレーオフはパリーグ 2 位の福岡ソフトバンクホークスがセリーグを勝ち上がった読売巨人軍をスweepし、日本一となって幕を閉じた。野球は投手力と言われることも多々あるが、近年は圧倒的打撃力を有する西武ライオンズがパリーグを制している。強打が売りで、犠打は 12 球団最小、かつ得点は 2 位に 100 点差以上つけており、従来の「スモールベースボール」と揶揄されていた日本型野球が影を潜めつつある。

日本型野球の代名詞ともいえる「送りバント」も近年その効果を疑問視され始めている。及川研・佐藤精一（2006）は送りバントを多用するチームでは、安打数の割には得点が上がらない傾向があり、得点期待値は送りバント無しのケースの方が大きくなることを分析している。

MLB では 2 番最強説なども流布し始め、2 番打者に強打者を置くチームが増え始めている。これは、送りバントや小技を 2 番打者に求める日本型野球とは相反する傾向である。セイバーメトリクスに基づく各打順の分析では、打席の多く回る上位全般が重要であり、2 番は併殺を回避できる脚力も求めることが多い。NPB でも MLB の流れを取り入れる球団がいくつかあり、2017 年の楽天は外国人のペゲーロを 2 番に据えたり、2019 年は数試合だけであるが DeNA の筒香が 2 番を務めたりしたこともあった。

このように、野球において今まで経験則で常識とされていたことが、セイバーメトリクスなどにより、覆され野球の試合運びが変わりつつある。その過渡期にある 2019 年度のデータを用いて得点への貢献度を見ていきたいと考えている。

2. 先行研究

鍋谷清治（2007）は 2001～2005 年のプロ野球公式戦と春夏の甲子園本戦の試合データを用いて、各打撃結果の理論分布への当てはめと 2 チーム間の和と差の分布について近似的な当てはめを行い、得点及び勝敗に対する回帰分析を行った。結果として、プロ・高校生を問わず、1 試合あたりの諸変数の分布の多くは通常为正または負の二項分布で近似可能であるが、2 チーム間の和と差の分布は近似的な理論分布を作ることではできなかった。また、得点及び勝敗に対する各変数の重要性はいずれも安打数がもっとも重要であり、犠打数は得点にあまり影響しない代わりに勝敗に対して大きな影響を与えるとわかった。

3. 分析

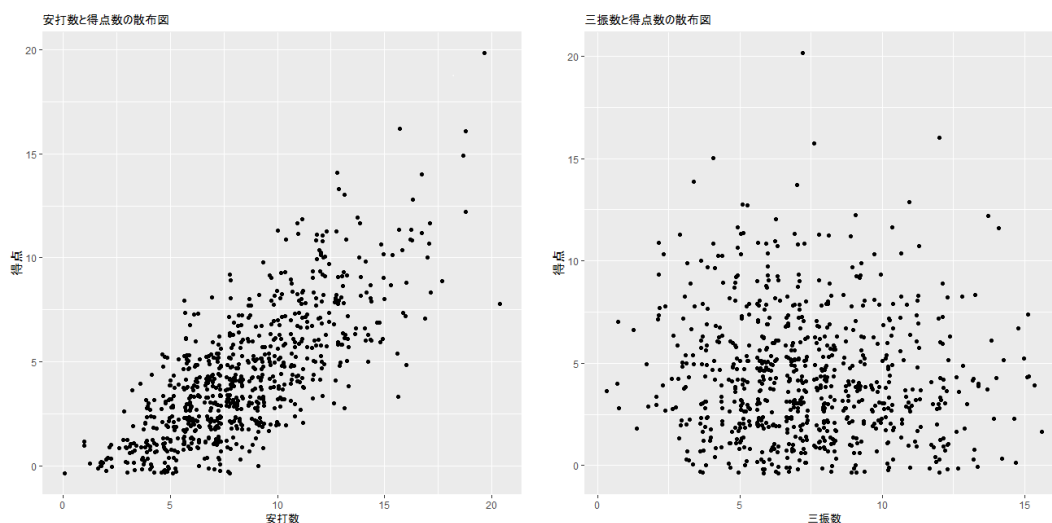
1) データ

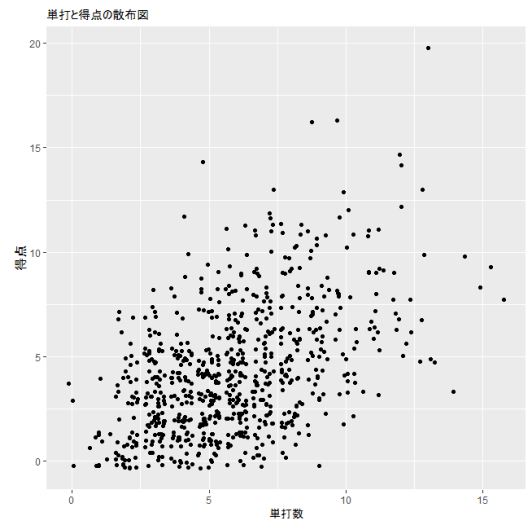
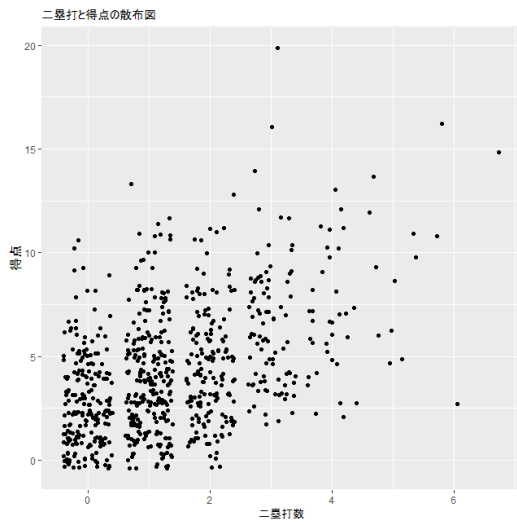
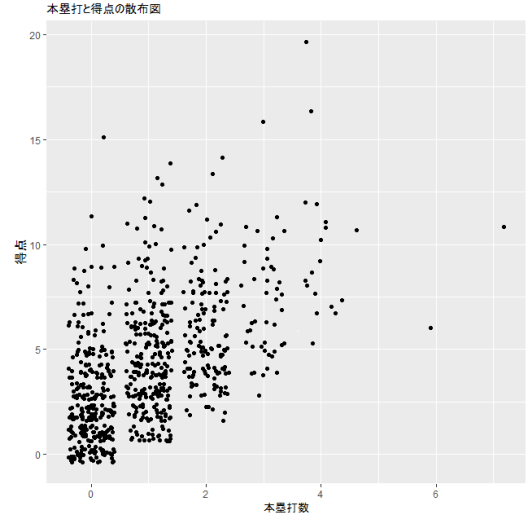
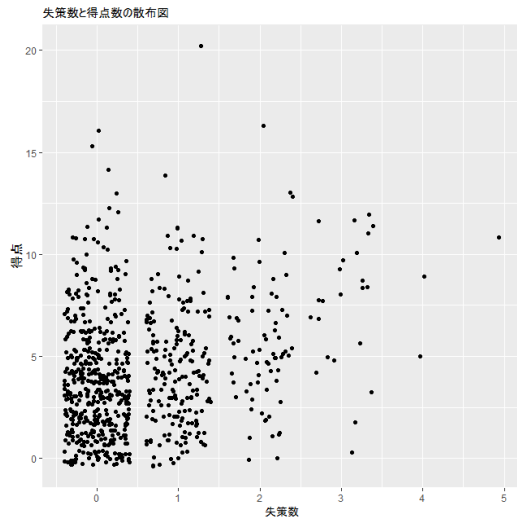
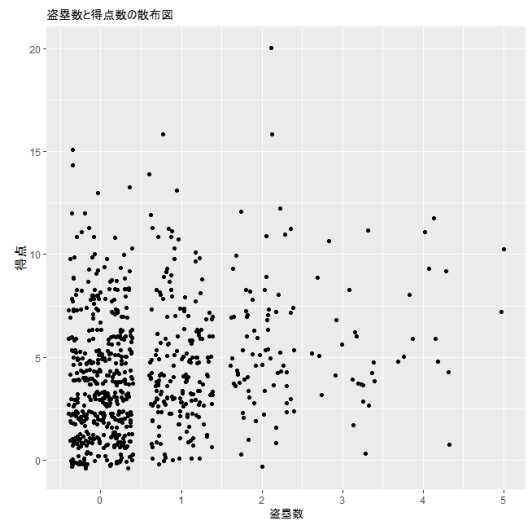
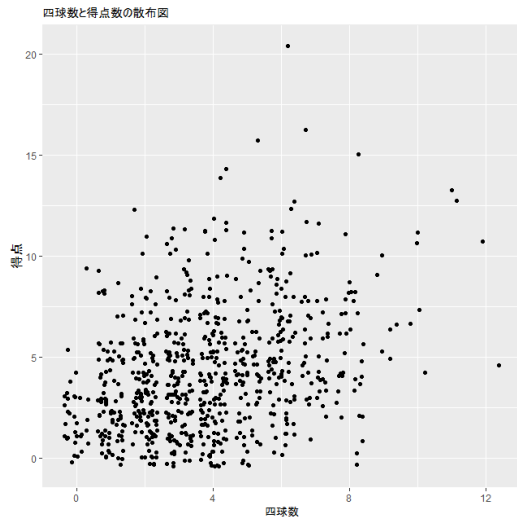
スポーツナビが出している 2019 年度のパリーグ 143 試合の結果をすべて入力し、打撃結果は以下の 9 つに分類した。

単打、二塁打、三塁打、本塁打、アウト、犠打、犠飛、併殺、四死球
FC は「失策」に含み、三振ゲッツー、タッチアップ失敗、ライナー飛び出しは「併殺」に含む。なお、今回の分析における失策とは、相手が犯した失策のことである。

2) 得点分析

重回帰分析を行う前に、各指標における得点との散布図と相関係数を求めていく。





安打数	三振数	四球数	盗塁数	失策数	本塁打数	二塁打数	単打数
0.753052	-0.08223	0.34321	0.223088	0.265441	0.538462	0.462545	0.489374

表 1 得点との相関係数

散布図に関しては、安打数や得点数はすべて整数値であり、同じ値をとりやすいことを踏まえ、多少ずらしてプロットし、密度がより分かりやすくなるように調整している。得点数と安打数は非常に強い相関があることが散布図からもうかがえる。また、三振数と得点は全くと言ってよいほど関係していないことがわかる。盗塁数と失策数、四球数は試合で起こる可能性が非常に低いものなので、0,1,2 の間に固まっており、相関はそこまで強いとは言えない。安打を本塁打、二塁打、単打に分けても、それぞれある程度の相関を持っていることが分かった。三塁打に関しては、本数が少なく、有意な結果は望めないため、今回は相関係数は求めなかった。まずは、安打を分けずに、得点を以下の式で回帰してみる。

$$y_{tokuten} = \alpha + \beta_h x_h + \beta_k x_k + \beta_b x_b + \beta_{st} x_{st} + \beta_{er} x_{er}$$

この時、各変数は以下のとおりである。

$y_{tokuten}$: 得点

x_h : 安打数

x_k : 三振数

x_b : 四球数

x_{st} : 盗塁数

x_{er} : 失策数

説明変数	係数	標準誤差	P 値	判定
安打数	0.63314	0.02041	2×10^{-16} 未満	***
三振数	-0.04422	0.02373	0.0628	.
四球数	0.32386	0.03071	2×10^{-16} 未満	***
盗塁数	0.02465	0.07452	0.7409	
失策数	0.44747	0.08214	6.94×10^{-8}	***

表 2 得点の回帰分析

決定係数 : 0.6471 調整済み : 0.6448

注意) 判定について、「***」は 0.1% 水準、「**」は 1% 水準、「*」は 5% 水準、「.」は 10% 水準である。

安打数、四球数、失策数に関しては、0.1%水準で有意であり、得点に大きな影響を与えているといえる。盗塁数に関しては、アウトカントを増やすことなく得点圏のチャンスを迎えることができ、得点に大きくつながりそうだが、数がそこまで多くなく、有意な結果は得られなかった。

ここで、安打を「本塁打」「三塁打」「二塁打」「一塁打」に分けて再度重回帰分析してみる。

$$y_{tokuten} = \alpha + \beta_{sh}x_{sh} + \beta_{dh}x_{dh} + \beta_{th}x_{th} + \beta_{hr}x_{hr} + \beta_b x_b + \beta_{er}x_{er} + \beta_k x_k + \beta_{st}x_{st}$$

$y_{tokuten}$: 得点

x_{sh} : 単打数

x_{dh} : 二塁打数

x_{th} : 三塁打数

x_{hr} : 本塁打数

x_b : 四球数

x_{er} : 失策数

x_k : 三振数

x_{st} : 盗塁数

説明変数	係数	標準誤差	P 値	判定
単打数	0.43577	0.02152	2×10^{-16} 未満	***
二塁打数	0.81078	0.04549	2×10^{-16} 未満	***
三塁打数	1.13821	0.13906	1.18×10^{-15}	***
本塁打数	1.38976	0.05319	2×10^{-16} 未満	***
四球数	0.30534	0.02597	2×10^{-16} 未満	***
失策数	0.47273	0.06942	2×10^{-16} 未満	***
三振数	-0.06148	0.02005	0.00224	**
盗塁数	0.04808	0.06281	0.44423	

表 3 安打を細分化した時の回帰分析

決定係数 : 0.7532 調整済み : 0.7504

注意)判定について、「***」は 0.1%水準、「**」は 1%水準、「*」は 5%水準である。

安打を詳細に分けると、決定係数が大幅に増加し、0.7504 となった。ほとんどの変数が 0.1%水準で有意となっており、三振数も 1%水準で有意なものとなった。俗に四球や失策は単打以上の重みがあると言われているが、係数も同じ位となった。四球に関しては、勝負を避けることを目的とした敬遠も存在しており、単打や失策よりは多少低くなることは容易に想像できる。また、失策は、投手の気落ちもさることながら、悪送球などはさらなる進塁も許すことがあり、結果的に二塁打と同じ結果をもたらすことなどもある為、多少大きくなっている。

単打と二塁打,三塁打の係数はほぼ 2 倍弱、3 倍弱となっており、塁打分係数が大きくなっているのは興味深い。一方、本塁打に関しては三塁打とさほど変わらない。三塁打自体がそこまで多くなく、脚力があるバッターが放つことが多く、得点につながりやすいのだろう。

なお、三振数については、アウトカウントが増えるのであるから、わずかでありながら、負の係数であることは想像に難くない。

2) 勝敗の分析

次に、勝敗に回帰してみる。

以下の式に関して、二項ロジスティック回帰で分析してみる。

$$\log \frac{p}{1-p} = \alpha + \beta_{sh}x_{sh} + \beta_{dh}x_{dh} + \beta_{th}x_{th} + \beta_{hr}x_{hr} + \beta_b x_b + \beta_{er}x_{er} + \beta_k x_k + \beta_{st}x_{st}$$

p : 勝利する確率

x_{sh} : 単打数

x_{dh} : 二塁打数

x_{th} : 三塁打数

x_{hr} : 本塁打数

x_b : 四球数

x_{er} : 失策数

x_k : 三振数

x_{st} : 盗塁数

説明変数	係数	標準誤差	P 値	判定
単打数	0.15597	0.03442	5.85×10^{-6}	***
二塁打数	0.31725	0.07258	1.24×10^{-5}	***
三塁打数	0.99795	0.23789	2.73×10^{-5}	***
本塁打数	0.50924	0.08801	7.20×10^{-9}	***
四球数	0.13556	0.04012	7.27×10^{-4}	***
失策数	0.39932	0.11371	4.45×10^{-4}	***
三振数	-0.14545	0.03186	5.00×10^{-6}	***
盗塁数	0.27099	0.10144	7.55×10^{-3}	**

表 4 勝敗のロジスティック回帰分析

AIC : 830.99

すべての変数に関して、1%水準で有意な影響を確認することができた。三振だけが負の影響を与え、それ以外は正の影響を与えていた。3 塁打の係数が非常に高かったが、本数自体が非常に少ないため、参考程度とした。モデルの説明力自体は低かったが、各項目は少なくとも影響を与えていることが分かった。勝敗に対しては、単打以上に失策が影響していることが分かった。また、四球もヒット 1 本に迫るくらい勝敗に影響を与えている。

3) 各打席に着目した得点分析

状況	1番	2番	3番	4番	5番	6番	7番	8番	9番	全体
無死無走者	41.3%	17.5%	17.3%	26.1%	24.3%	22.2%	22.5%	23.9%	23.9%	24.4%
無死1塁	4.6%	11.4%	4.7%	5.3%	6.6%	6.2%	5.3%	5.1%	5.2%	6.1%
無死2塁	0.7%	1.8%	1.3%	1.1%	1.2%	1.2%	1.3%	1.0%	1.1%	1.2%
無死3塁	0.0%	0.5%	0.2%	0.3%	0.0%	0.1%	0.1%	0.1%	0.0%	0.1%
無死12塁	0.6%	0.8%	1.9%	1.3%	1.4%	2.2%	1.4%	1.2%	0.9%	1.3%
無死13塁	0.3%	0.3%	0.8%	0.5%	0.5%	0.4%	0.4%	0.5%	0.3%	0.5%
無死23塁	0.0%	0.1%	0.1%	0.3%	0.3%	0.3%	0.3%	0.3%	0.0%	0.2%
無死満塁	0.1%	47.6%	0.2%	32.5%	0.2%	26.4%	0.7%	35.6%	0.4%	34.7%
1死無走者	14.3%	28.2%	13.3%	12.0%	18.7%	17.6%	16.5%	17.3%	18.8%	17.4%
1死1塁	5.4%	6.3%	10.7%	6.2%	5.5%	8.7%	7.4%	6.5%	5.8%	7.0%
1死2塁	2.9%	2.6%	4.8%	3.0%	1.7%	2.4%	2.4%	2.4%	3.3%	2.8%
1死3塁	0.6%	0.6%	1.8%	0.7%	0.4%	0.7%	0.7%	0.8%	0.8%	0.8%
1死12塁	1.5%	2.3%	2.2%	3.7%	2.9%	2.8%	3.6%	2.9%	2.1%	2.6%
1死13塁	1.0%	1.2%	0.8%	1.9%	1.5%	0.7%	1.3%	0.9%	1.1%	1.2%
1死23塁	0.4%	0.4%	0.6%	0.8%	0.5%	0.7%	0.7%	0.7%	1.2%	0.7%
1死満塁	0.7%	26.9%	0.7%	42.3%	0.8%	34.9%	0.8%	29.1%	1.6%	32.8%
2死無走者	11.6%	10.5%	22.8%	11.2%	11.2%	14.7%	14.6%	13.3%	14.4%	13.8%
2死1塁	4.1%	5.5%	5.6%	11.3%	6.2%	5.9%	8.1%	7.4%	6.5%	6.7%
2死2塁	3.2%	2.9%	2.9%	4.1%	3.0%	2.4%	2.9%	3.4%	3.4%	3.1%
2死3塁	0.9%	1.2%	1.5%	2.2%	1.4%	0.7%	1.6%	1.4%	1.0%	1.3%
2死12塁	2.2%	2.2%	2.6%	3.1%	5.9%	4.1%	3.4%	4.2%	4.2%	3.5%
2死13塁	1.4%	1.3%	1.7%	1.5%	2.6%	1.9%	1.6%	2.0%	1.9%	1.7%
2死23塁	0.7%	0.5%	0.7%	0.3%	0.9%	0.8%	0.9%	1.0%	0.8%	0.7%
2死満塁	1.4%	25.6%	1.1%	25.2%	0.9%	38.7%	1.5%	35.3%	1.3%	32.5%
合計	25.6%	1.1%	25.2%	0.9%	38.7%	1.5%	35.3%	1.3%	32.5%	1.8%
全体	34.1%	33.5%	33.5%	33.5%	33.5%	33.5%	33.5%	33.5%	33.5%	33.5%

表 5 シチュエーション別の打席数

冒頭で述べた通り、「二番最強説」などにも興味があり、各打順を区別して各打席での結果を集計した。各打順のシチュエーション割合は以下の表のとおりとなった。

1番打者の無死無走者が多いのは、初回の攻撃は100%1番から始まるため、理解しやすい。1番2番は2死以外で打順が回ってくることが多く、チャンスメイクするなど、大きな役割を果たしている。3番4番は得点圏での打撃が多く、特に4番は2死でのチャンスが多いため、一振りで決めることが求められる。

次に各打順の得点貢献度を見ていきたい。今回、各打者が打席に立った後にそのイニングでとった点数を得点としてカウントした。これは、出塁や進塁打がその後の得点に影響することを踏まえてこの方式をとった。例えば、以下のような打者一巡の攻撃の場合、得点の記録は一番右の値を記録している。

	状況	打撃結果	得点
1番	無死無走者	本塁打①	3
2番	無死無走者	遊飛	2
3番	1死無走者	二安	2
4番	1死1塁	左中二①	2
5番	1死2塁	中安	1
6番	1死13塁	四球	1
7番	1死満塁	左犠飛①	1
8番	2死12塁	三失	0
9番	2死満塁	三振	0

表 6 今回の得点の定義

まず、無死無走者の場合の得点期待値及び得点率を見てみる。得点率は点を取ることができるかどうか、得点期待値は攻撃の質まで見ることができる。この場合、先頭バッターがどの打順から始まる時に点が入りやすいか、また、出塁時はどれだけ得点に貢献しているのかがわかる。野球ではよく先頭打者が肝心と言われているが、どれだけ重要なのか見ていきたい。

総じて先頭打者が出塁すると、5割近くの確率で得点できることが分かった。特に2,3番から始まるイニングは得点率が高く、期待値も高かつ

た。意外だったのは、9番打者の出塁時得点期待値が非常に高かったことだ。出塁率自体は低いですが、出塁すると上位に回っていくこともあり、得点が増える傾向にあるのだろう。逆に、4、5番打者から始まる攻撃は出塁しても下位に下っていくことになるので、得点率も期待値も低いものとなっていた。

	得点率	得点率(出塁時)	得点期待値	得点期待値(出塁時)
1番	29.4%	50.4%	0.54	0.98
2番	29.7%	51.8%	0.57	1.04
3番	30.6%	54.3%	0.55	1.01
4番	26.9%	48.0%	0.44	0.84
5番	24.0%	46.1%	0.41	0.82
6番	24.8%	50.4%	0.42	0.90
7番	23.1%	48.2%	0.41	0.89
8番	20.9%	44.7%	0.39	0.89
9番	27.5%	53.6%	0.54	1.11
総合	26.5%	49.7%	0.48	0.94

表 7 打順別の得点貢献度

最後に各打席結果を踏まえた、得点要因の重回帰分析を行ってみた。打席数はリーグ 429 試合の合計 33325 打席分を対象とした。得点に関しては、整数値をとり、そこまで頻度も高くないため、ポアソン分布に回帰することにした。

打順に関しては、2,6,7,8番が有意であったので、その打順のみで回帰した。

$Y_{tokuten} \sim \text{Poisson}$

$$= \alpha + \beta_1 D_2 + \beta_2 D_6 + \beta_3 D_7 + \beta_4 D_8 + \beta_5 D_{sh} + \beta_6 D_{ah} + \beta_7 D_{th} + \beta_8 D_{hr} + \beta_9 D_o + \beta_{10} D_{do} + \beta_{11} D_{bb} + \beta_{12} D_{er} + \beta_{13} D_{sb}$$

今回の得点は上述の通り、各打席後にそのイニングで何点入ったかをデータとしている。ダミー変数は順に「2番打者」「6番打者」「7番打者」「8番打者」「単打」「二塁打」「三塁打」「本塁打」「アウト」「併殺」「四死球」

「(相手の) 失策」「犠打」とした。

説明変数	係数	標準誤差	P 値	判定
2 番打者	0.09050	0.02449	2.20×10^{-4}	***
6 番打者	-0.13011	0.02765	2.53×10^{-6}	***
7 番打者	-0.15191	0.02842	9.07×10^{-8}	***
8 番打者	-0.15594	0.02947	1.22×10^{-7}	***
単打	-0.41180	0.06121	1.72×10^{-11}	***
二塁打	-0.09944	0.06465	1.24×10^{-1}	
三塁打	0.24207	0.08710	5.45×10^{-3}	**
本塁打	0.24525	0.06445	1.42×10^{-4}	***
アウト	-2.00375	0.06158	2×10^{-16} 未満	***
四死球	-0.60047	0.06268	2×10^{-16} 未満	***
併殺	-2.88199	0.14738	2×10^{-16} 未満	***
失策	-0.20692	0.07799	7.97×10^{-3}	**
犠打	-0.44867	0.07545	2.74×10^{-9}	***

表 8 得点をポアソン分布に回帰した重回帰分析

二塁打をのぞくすべてのダミー変数が有意な影響を与えていることが分かった。2 番打者はわずかではあるが、得点に対して正の係数を持っており、他の打順では負の係数か、もしくは有意な結果を得ることができなかった。打撃結果に関わらず、打順が回ってくるというだけで正の係数をわずかながら持っているということは、打順の中でも重要な場所にあることがわかる。6,7,8 番は俗にいう下位打線であり、上位にもつながっていきやすいことを踏まえると、得点要因にはなりにくいのだろう。単打に関しては、割と出やすいわりに、得点に絡まないことが多いため、負の係数となった。1 イニングで 3 本単打が出ても、無得点に終わることがあり、また次打者の併殺などの危険性もはらんでいるため、このような値となったのであろう。3 塁打と本塁打に関しては、同様に正の影響を与えている。本塁打は必ず得点につながり、3 塁打も前述の通り得点につながりやすい要素が詰まっているため、このような結果となった。

一方、アウトと併殺は非常に大きく負の係数をとった。これは 1 イニングで 3 アウトとると、イニングが終わってしまう野球のルールを考えると、

当然の結果といえる。犠打に関しては、犠飛も含んでいるにもかかわらず、単打よりも低い結果となった。犠打は得点圏を作り出しているはずにもかかわらず、得点に対して、そこまで影響がないということが分かった。

ただし、今回の分析は各シチュエーションでのデータが少なく、すべてまとめて一括で分析した。単打一本でも、無死二塁の一本と二死無走者の一本とでは重みが全く異なる。今後はそこも分けて考えると、もっとモデルの説明力がついてくると考えられる。

4. 考察・まとめ

今回の分析において、まずは、2019 年度プロ野球パリーグ公式戦の試合ごとの安打等の数を集計し、得点及び勝敗への影響を回帰分析した。四死球や失策は単打と同じくらい得点に影響し、二塁打や三塁打はより大きく影響を与えていた。勝敗に対しても各安打は正の影響を与え、三振数のみ負の影響を与えていた。ただし、モデルの説明力は高くなく、今後考えられる変数は「味方の投手力」や「ビジターor ホーム」などが考えられる。

次に、打席 1 つ 1 つに着目し、得点への貢献を考察した。打順やシチュエーションなど、細分化したかったが、1 リーグ、1 年分のデータしか集めることができなかつたため、なかなか踏み込んだ分析はできなかつた。ただ、それぞれの結果をダミー変数として組むと、ある程度は予想できる結果となっており、それぞれの変数自体の影響力はわずかながら確認することができた。2 番打者はわずかながら、正の影響を与えていたので、2 番最強説について、今後もさらにデータを集め、さらに最適なモデルを考案できればさらに面白くなっていくと考える。

5. 参考文献

◆ 及川研・佐藤精一 (2006)

「送りバントをした場合としない場合の得点期待値の差異について～日本のプロ野球公式戦を対象として」、東京学芸大学紀要 芸術・スポーツ科学系 58、p.123-128

◆ 鍋谷清治 (2007)

「野球のデータの統計的分析」、日本統計学会誌.シリーズ J 36(2)、p91-115

◆ スポーツナビ

<https://sports.yahoo.co.jp/>