

順序ロジットモデルでプロ野球分析

著者	根岸 紳
雑誌名	産研論集
号	47
ページ	65-68
発行年	2020-03-20
URL	http://hdl.handle.net/10236/00028665

順序ロジットモデルでプロ野球分析

根 岸 紳

1. はじめに

2019年のプロ野球シーズン中、阪神タイガースについて、しばしば、新人である近本光司選手の活躍がチームに元気を与え、チームの勝利に貢献しているといわれていた¹⁾。近本選手が、ヒットであれ死四球であれ、出塁すると後続の選手の活躍を生み、上手くいけば得点につながるということであるのだろうか。本稿ではこのことがデータで検証できるのかどうかを探りたい。近本選手は1番打者であり、外野手であるので、リーグは異なるがロッテマリーンズ1番打者で外野手の荻野貴司選手と比較しながら検討したい。また、阪神タイガースで近本選手の後続として、糸原健斗内野手を取り上げる。糸原選手はシーズン中、近本選手に続く2番打者として多くの試合に出場していた。計量分析の方法は順序ロジットモデルを採用する。また、データはYahooのSportsnavi「プロ野球」の中にある毎試合の記録を利用した²⁾。

2. 基礎データとマクロ分析

近本光司選手と荻野貴司選手は、時期は異なるが同じ関西学院大学硬式野球部の出身であり、ともに2年間社会人野球を経験し、その後プロ野球の世界に入ってきた。近本選手は1年目の新人、荻野選手は10年目のベテランである。近本選手はセ・リーグ、荻野選手はパ・リーグに属し、どちらも1番バッターであり³⁾、ほとんどともにセンター（中堅）を守っている。ふたりの成績データを分析しながら、近本選手のチームへの貢献についてその特徴を検出したい。また、同じ球団内

の比較も必要であろう。そのため2番で打つことが多かったプロ野球3年目の糸原内野手を取り上げる。ただ、糸原選手は、7月末あたりから2番ではなく7番あたりを打つことが多くなった。なお、糸原選手も、大学のあと、社会人野球を2年間経験している。

3選手の身長、体重、打撃の成績等について、(表1)で見てみよう。3選手ともあまり大きくない体格であるが、近本選手と荻野選手はよく似た体格である。試合数については、近本選手は1試合のみ欠場、糸原選手は全試合に出場し、荻野選手は、腰痛のため登録抹消の期間があり、近本、糸原選手より試合数は少ない。糸原選手と比較して近本、荻野両選手に共通するのは、盗塁の多さと本塁打の本数であり、近本選手はセ・リーグの盗塁王となる。打数の多い近本選手は三振も多く、相対的

表1 三選手の基礎データ

2019年シーズン	近本光司	荻野貴司	糸原健斗
身長	170cm	172cm	175cm
体重	72kg	75kg	77kg
生年月日	1994.11.9	1985.10.21	1992.11.11
投打	左	右	右投げ左打ち
出場試合数	142	125	143
打数	586	508	491
得点	81	76	66
安打	159	160	131
打点	42	46	45
三振	110	56	70
四死球	37	48	67
犠打	14	9	11
盗塁	36	30	6
失策	4	1	7
本塁打	9	10	2

(出所) Yahoo.Sportsnavi、日本野球機構 NPB.jp

- 1) 例えば、2019年10月3日朝日新聞夕刊「虎バン主義。」の中で高野純一ABCアナウンサーは次のように発言している。『今年には近本の年。新人で盗塁王も獲得しましたし、出塁すれば後続の打者に好影響を及ぼすことも証明してきました。』
- 2) 例えば2019年9月1日阪神タイガース対読売ジャイアンツのデータは <https://baseball.yahoo.co.jp/npb/game/2019090102/stats> を見ることによって得られます。
- 3) ただし、回数は少ないが、近本選手は1番以外を打つこともあった。

に打数が少ない糸原選手に四死球が多い。荻野選手は、打率が最終的にパ・リーグで4位となり、打率3割を超えるバッターとなる。

一般に打者成績の主要なものとして、「打率」「本塁打」「打点」の3つがあげられることが多い。それぞれ、首位打者、本塁打王、打点王と呼ばれ、表彰の対象となる。この3つのタイトルを独占した選手は各リーグで三冠王として最高の賞賛を受ける。しかし、これらは必ずしも選手の「得点獲得能力」を表したものではないといわれ、長打率と出塁率を合計したOPS (On-base Plus Slugging) が得点との相関が高いということで、OPSは重要視されるようになってきている⁴⁾。日本野球機構のオフィシャルサイト⁵⁾には個人別成績があり、長打率と出塁率が公表されている。また、得点能力には、走力、さらに犠打といったランナーを進塁させる能力を考慮した打撃指標も提案されている⁶⁾。これらから学ぶことは、打者の能力にとって、打率、打点、本塁打以外に四球、死球、犠打、盗塁も大事であることである。

打率は安打数を打数で割って求めるが、OPS⁷⁾は出塁率と長打率を合計して求める。3選手の打率とOPSの値をまとめてみよう。

近本光司 打率0.271 OPS 0.689= 長打率0.375 + 出塁率0.313
 荻野貴司 打率0.315 OPS 0.842= 長打率0.470 + 出塁率0.371
 糸原健斗 打率0.267 OPS 0.689= 長打率

0.336 + 出塁率0.353

OPSの値はいくらぐらいが評価の対象になるのだろうか。鳥越他(2014)によれば、2013年シーズンのOPSランキングが掲載されているが、例えばセ・リーグの8位は阪神のマーティンの0.845、パ・リーグの4位はロッテの井口資仁の0.902という高い数値である。そうなると、荻野選手のOPSは高く、近本、糸原両選手のOPSはあまり高くないことになる。

マクロ的に見ると、打率もOPSも荻野選手が大きくリードしているといえる。以上はマクロ分析である。次に、ミクロ的に見て行こう。

3. ミクロ分析

以下の方法はマイクロデータをもとにしたマイクロ計量分析である。ソフトウェアはExcelとgretl⁸⁾を使った。

各選手の得点、安打、打点、犠打、盗塁などの成績の中でなにがチームの勝敗に貢献しているのであろうか。これを推定、検定するために、以下の被説明変数と説明変数を設定した。

被説明変数 Y: 0=負け、1=引き分け、2=勝ち
 説明変数 X1: 打数、X2: 得点、X3: 安打、X4: 打点、X5: 三振、X6: 四死球、X7: 犠打、X8: 盗塁、X9: 失策、X10: 本塁打

まず、被説明変数、説明変数の相関係数を調べてみよう。その結果は(表2)、(表3)、(表4)の通りである。

表2 近本光司外野手の相関行列

近本光司	勝敗	打数	得点	安打	打点	三振	四死球	犠打	盗塁	失策	本塁打
勝敗	1										
打数	0.040	1									
得点	0.344	-0.047	1								
安打	0.222	0.188	0.533	1							
打点	0.262	0.030	0.237	0.330	1						
三振	0.027	0.276	-0.195	-0.279	-0.081	1					
四死球	0.116	-0.326	0.356	0.133	0.083	-0.115	1				
犠打	0.118	-0.186	-0.133	-0.087	-0.004	-0.077	-0.158	1			
盗塁	0.226	-0.136	0.262	0.181	0.124	-0.188	0.317	-0.064	1		
失策	-0.175	-0.096	-0.077	-0.108	-0.077	-0.004	-0.003	-0.056	0.075	1	
本塁打	0.086	-0.078	0.327	0.239	0.506	-0.100	0.089	0.012	-0.065	-0.044	1

4) 鳥越他(2014)「得点能力」を評価する指標OPS」(pp.28-31)

5) <http://npb.jp/bis/players>

6) 鳥越他(2014)「走塁能力や進塁させる能力も加味した打撃指標」(pp.31-33)

7) OPSを計算するときに使う出塁率と長打率の定義は鳥越他(2014)を参照せよ。

8) 順序ロジットモデルの推計にはgretlを使った。使い方については加藤(2012)が詳しい。

順序ロジットモデルでプロ野球分析

説明変数間の相関は、近本選手と荻野選手は似ているところが多い。例えば、相関係数の相対的に高いところでは、安打と得点の相関、本塁打と打点の相関が類似している。次に、糸原選手の相関行列表を見てみよう。

糸原選手の説明変数間の相関については、安打と得点の相関が相対的に高く、これは近本、荻野両選手も高い。被説明変数と説明変数の相関については、3選手の中で、近本選手は複数の説明変数と相対的に相関が高いのが特徴的である。

それでは、(0、1、2) からなる被説明変数、そ

して10個の説明変数からなる離散選択モデルを計測しよう。被説明変数が連続変数ではなく、離散変数をとる場合、離散選択モデルと呼ばれている⁹⁾。離散選択モデルの中で、順序ロジットモデルとは被説明変数のとりうる値(選択肢)が3つ以上あり、かつ、それらに何らかの順序がある場合に適用するものである。「勝つ」、「引き分け」、「負ける」には順序があると想定し、「2」、「1」、「0」とあてはめた。説明変数は打数から本塁打まで10種類があり、これらのデータをもとに、順序ロジットモデルを推計した。この結果は、(表5)にまと

表3 荻野貴司外野手の相関行列

荻野貴司	勝敗	打数	得点	安打	打点	三振	四死球	犠打	盗塁	失策	本塁打
勝敗	1										
打数	-0.121	1									
得点	0.335	0.015	1								
安打	0.110	0.328	0.556	1							
打点	0.099	0.267	0.226	0.346	1						
三振	0.000	0.168	-0.037	-0.115	-0.046	1					
四死球	0.164	-0.336	0.253	0.034	-0.063	-0.053	1				
犠打	0.125	-0.191	-0.020	-0.048	-0.055	-0.052	0.021	1			
盗塁	0.118	0.042	0.155	0.248	-0.069	0.070	0.309	0.054	1		
失策	-0.091	0.087	-0.075	-0.025	-0.045	-0.066	-0.061	-0.025	-0.045	1	
本塁打	0.108	0.117	0.365	0.262	0.559	-0.108	0.000	-0.074	-0.134	-0.024	1

表4 糸原健斗内野手の相関行列

糸原健斗	勝敗	打数	得点	安打	打点	三振	四死球	犠打	盗塁	失策	本塁打
勝敗	1										
打数	-0.010	1									
得点	0.298	0.195	1								
安打	0.248	0.288	0.476	1							
打点	0.189	0.080	0.160	0.367	1						
三振	-0.122	0.206	-0.069	-0.133	-0.163	1					
四死球	0.206	-0.300	0.191	0.048	0.097	-0.148	1				
犠打	0.023	-0.064	-0.070	0.059	-0.094	-0.054	-0.076	1			
盗塁	0.070	0.047	0.110	0.112	-0.049	0.004	-0.093	0.184	1		
失策	0.098	0.093	-0.056	0.108	-0.010	-0.023	-0.061	0.051	-0.047	1	
本塁打	0.121	0.065	0.259	0.166	0.318	-0.095	0.006	-0.031	-0.025	-0.027	1

表5 近本、荻野、糸原3選手の計測結果

被説明変数	順序ロジットモデル		
	勝敗	勝ち=2、引き分け=1、負け=0	
説明変数	近本光司	荻野貴司	糸原健斗
打数	0.164 (0.766)	-0.237 (-0.977)	-0.093 (-0.509)
得点	1.225 (3.139)***	1.137 (2.999)***	0.740 (2.174)**
安打	-0.018 (-0.069)	-0.237 (-0.910)	0.236 (0.811)
打点	1.220 (2.482)**	0.299 (0.929)	0.415 (1.269)
三振	0.380 (1.556)	0.126 (0.383)	-0.180 (-0.593)
四死球	0.068 (0.155)	0.094 (0.255)	0.476 (1.622)
犠打	1.464 (2.260)**	0.937 (1.216)	0.219 (0.352)
盗塁	0.876 (1.897)*	0.286 (0.644)	0.849 (0.846)
失策	-16.54 (-0.008)	-10.29 (-0.035)	1.201 (1.358)
本塁打	-1.640 (-1.537)	-0.130 (-0.155)	9.768 (0.040)
対数尤度	-97.80	-88.04	-106.85
サンプルサイズ	142	125	143

9) 山本 (2015) 第8章「順序ロジットモデルと多項ロジットモデル—離散選択モデルの応用—」

められている。

荻野選手、糸原選手は X2 の得点だけが、被説明変数 Y の勝敗に有意に影響を与えている。一方、近本選手は得点のほかに、X4 の打点、X7 の犠打、X8 の盗塁というように 4 つの要因が勝敗に有意にプラスの影響を与えている。近本選手と荻野選手の比較から、マクロ的には圧倒的に荻野選手が打者として上位にあるけれども、ミクロ的な分析から、近本選手のほうが荻野選手に比べて多くの要因によってチームの勝敗に貢献している可能性がうかがえる。

なお、順序ロジットモデルは、閾値 (cut-point) も計算できる¹⁰⁾。結果は表示していないが、閾値の推定からも近本選手の結果は統計的に有意に検定されていることから、近本選手の活躍は勝ち負け、引き分けがはっきり分類できる打撃成績であったことがうかがえる。荻野選手、糸原選手の推計では、有意な閾値は得られていない。

4. 結びにかえて

本稿で検討したモデルは、被説明変数が質的な数値で説明変数が量的な数値であった。チームの勝利は「2」という値であったが何点を入れて勝ったのか、引き分けは「1」であったが何点で引き分けであったのか、負けは「0」であったが何点で負けたのかの数値データで分析することも可能であろう。したがって、われわれの計測結果は被説明変数を質的に考えた限定的な結果であるといえる。

2019年8月7日神宮球場で行われたヤクルト阪神戦では阪神は11対2で負けたが、このとき、デイリースポーツ online¹¹⁾ は次のように近本選手の言葉を伝えている。『 』内が近本選手の言葉である。

近本選手は「懸命に走った。見事に2盗塁を決め、最後まで諦めない姿勢は貫いた。『試合の展開的にも差がついてしまったところで走ったんですけど、もっと序盤だったり、初回というところで自分の仕事ができればよかった。』 常々、話すのは

『ホームにかえってくるのが自分の仕事なので』と本塁に生還することを強く意識している。』

われわれの計測結果もこのことを裏付けている。得点がチームの勝利に貢献していることは近本選手をはじめ荻野選手、糸原選手とともに検証されているが、近本選手にはそのほか盗塁も勝利に貢献している。さらに近本選手の犠打や打点もチームの勝利に貢献していることも検証できた。打者にとって、打率や OPS のようなマクロ的な数量では秀でていなくとも、1試合ごとのミクロ的な分析によればマクロでは見えない所属チームの勝利への貢献がありうることを示すことができた。

参考文献

- 加藤久和 (2012) 『gretl で計量経済分析』、日本評論社
鳥越規央・データスタジアム野球事業部 (2014) 『勝てる野球の統計学 セイバーメトリクス』、岩波科学ライブラリー 223、岩波書店
山本勲 (2015) 『実証分析のための計量経済学 正しい手法と結果の読み方』、中央経済社

10) 閾値については、山本 (2015) の第 8 章「順序ロジットモデルと多項ロジットモデル」の中の「順序ロジットモデル」(pp.121-22) を参照せよ。

11) <https://www.daily.co.jp>