

プロ野球における投球系列データからのパターン抽出

守屋 拓馬^{*1}, 川又 泰介^{*2}, 松田 源立^{*3}

Pattern Extraction from Pitch Sequences in Professional Baseball

Takuma Moriya^{*1}, Taisuke Kawamata^{*2}, Yoshitatsu Matsuda^{*3}

ABSTRACT : In baseball, the pitch sequence is a significant factor in determining the outcome of a game. However, previous research has primarily focused on the direct relation between outcome and pitches rather than pitch sequences. In order to improve the game strategy in baseball, this study investigates the difference between the pitch sequences in outs and those in hits. For this purpose, pitching data of six pitchers of Nippon Professional Baseball (NPB) in 2022 was collected. Sequential pattern mining was then applied to extract common patterns in the pitch sequences. The results showed that the accuracy of the final phase pitches is crucial for achieving an out and that a pitch sequence incorporating fast-breaking is effective. Furthermore, the extracted patterns in hit suggested the appropriate inning to change pitchers.

Keywords : Sports analytics, Baseball statistics, Pitch sequence, Sequential pattern mining

(Received Sep 4, 2023)

1. はじめに

現代のプロ野球では投手のレベルが非常に上がっている。近年は投手の成績が上がり、打者の成績が下がる、いわゆる投高打低の時代と言われ、特に今年は完全試合達成者が1人、ノーヒットノーラン達成者が4人も現れた。その背景としては、奪三振能力の高い投手が多く現れていることが挙げられる。三振とは打者がストライクを3つ取られた場合に成立するものであり、ピンチの際にランナーを進めさせず、状況を維持した状態で確実にアウトカウントを1つ重ねることができる。このことから、失点をしないためには三振を取ることが大切であり、その配球術は各球団のスコアラーに日々研究されている。

また、現在公開されているプロ野球のデータとして、ゾーン別被打率や、カウント別被打率、球種別被打率などの状況別被打率が挙げられる。ゾーン別被打率は各ゾーンの結果球を収集し、それらのアウトにとった球数と

ヒットを打たれた球数で打率を計算する。カウント別、球種別でも同様に計算を行う。これらのデータにより、各投手の得意なゾーンや得意な球種などを分析することが可能になった。しかし、これらの分析は全て結果球による分析であり、その結果球に至るまでに投げられた投球に着目した配球に関するデータは、各球団のスコアラーが研究はしているものの公開はされていない。

本研究では結果球だけでなく、その前に投げられた球も含めた投球の流れに着目し、打者をアウトにとった配球とヒットを打たれた配球にどのような違いがあるかを調査する。データにはプロ野球一球速報データから6投手の投球をWebスクレイピングで収集したものを用いた。また、それらのデータからシーケンシャルパターンマイニングによるパターン抽出を行うことでアウトにとった配球とヒットを打たれた配球のそれぞれにどのような特徴があるのかを調査する。

2. 先行研究

2.1 シーケンシャルパターンマイニングを用いた麻雀の捨て牌の傾向分析

[1]ではオンライン麻雀の捨て牌の系列記録を収集し、

*1 : 理工学部情報科学科学部学生 (2023年3月卒業)

*2 : 理工学部情報科学科助教
(2023年4月より茨城大学情報戦略機構助教)

*3 : 理工学部情報科学科 (理工学科) 准教授
matsuda@st.seikei.ac.jp

そのデータ列から頻出パターン系列を、シーケンシャルパターンマイニングを用いて抽出することで、勝者と敗者の打牌パターンに違いがあるかを調査していた。結果、敗者には安全牌の後に危険牌を捨てるというパターンが現れ、勝者にはこのパターンが現れていないことが分かった。このことから麻雀は守りながら戦うことの方が有効であると結論付けていた。系列記録を収集し、そのデータ列からシーケンシャルパターンマイニングを用いてパターンを抽出するという点で本研究と類似しており、分析手法を参考にした。データ収集の対象を麻雀ではなく野球にしているという点が本研究との違いである。

2.2 最終球への配球推移に基づくキャッチャーの成績分析

[2]では捕手の性質を配球傾向から調査していた。主成分分析やベクトルなどを用いて全体の配球傾向と各捕手の配球傾向の違いを調査していた。結果としては、配球傾向が他の選手と異なる捕手ほど、防御率が低く好成績であり、逆に防御率の高い捕手は配球傾向が全体と類似していることが多いという結果が得られた。本研究では、配球研究の手法や考え方について参考にした。また、本研究との違いは研究対象が捕手であるか、投手であるかという点である。

2.3 プロ野球全球データから見る統計分析と実用性

[3]では、野球のスコアラーが記録した全球データを使用し、打者と投手の相性を分析するという行っていた。各投手にストライクを取れた球種とコース上位3種を投手の得意球とし、各打者にヒットが生まれた球種とコース上位3種を打者の得意球とし、この得意球に重みを付けて相性の判別を行った。結果として、得意球による相性の判別と実際の相性が一致するとは言えず、相性には球種やコース以外の要因があるということが分かった。収集した球種やコースのデータをどのような研究に活かすことができるかという点で参考にした。また、論文が発表されたのが2019年であったため、本研究では最新の傾向も探ることが出来ればよいと考えた。

2.4 大学野球における配球について：カウント0-0における投球の分析

[4]では大学野球のゲームにおいて最も多い右投手と右打者の組み合わせにおいて、カウント0-0からの投球についての現状を分析し、どのような投球がストライクを先取する上で有利になるのかを調査していた。全体としてストレートやスライダーが多いことや、打者がスイ

ングした場合に打ち損じている可能性が高いことからストライクゾーンの勝負をするべきであること、見送りによるストライクを取得するにはカーブやスライダーが有効であることなどが結果として示されていた。研究の分析手法やその考え方について参考にした。本研究との違いはカウントに着目した研究であるか、打席の投球の流れ全体に着目した研究であるかという点である。

3. 研究手法

3.1 データ収集

Pythonを用いて、Yahoo!が運営するスポーツナビ[5]からWebスクレイピングを行い、6投手の2022年のプロ野球で投げられた約17,000球、2,700打席のデータを収集した。

打席内の投球情報として、球種、高さ、コースの3種類の属性を収集した。SPMF[6]を適用するために投球ごとに-1、打席ごとに-2を格納し、さらに取り出した属性を英語に変換した。SPMFに適用するうえで-1はアイテムセットの終わり、-2はアイテムセットの集合であるパターン終わりを指すため、本研究では-1を投球ごと、-2を打席ごとに記述した。球種、高さ、コースの設定した各属性とその意味を表をTable. 1、Table. 2、Table. 3で示す。表の球種属性、高さ属性、コース属性の中からそ

Table. 1. 球種属性とその意味

straight	ストレート
fbrake	投手のストレート最高速度との差が20km/h未満の変化球
sbrake	投手のストレート最高速度との差が20km/h以上の変化球

Table. 2. 高さ属性とその意味

highb	高めボール球
highs	高めストライク
middleh	高さ真ん中
lows	低めストライク
lowb	低めボール球

Table. 3. コース属性とその意味

inb	インコースボール球
ins	インコースストライク
middlec	コース真ん中
outs	アウトコースストライク
outb	アウトコースボール球

それぞれ1つずつ選択し、データの収集を行った。また、[7]を参考に、BeautifulSoupを用いたPythonのプログラムを作成し、特定のタグ内に格納されている球種情報やコースと高さの座標を収集することで、各属性の要素を収集した。収集されたデータの一部をFig. 1 に示す。6 投手分それぞれ、アウトにとった打席の投球データと、ヒットを打たれた打席の投球データを記録し、計 12 個の投球データ兼入力データを作成した。また、6 投手全ての投球が記録されたデータもアウトとヒットそれぞれで作成し、計 14 個のデータの収集を行った。

3. 2 SPMFによる分析

投球データのパターン抽出のため、本研究ではSPMF [6]というソフトを使用した。SPMFはシーケンシャルパターンマイニングを行うソフトであり、今回はその1つである、「PrefixSpan_with_Strings」をアルゴリズムとして使用した。入力データはデータ収集で収集した 14 個のファイルを使用し、これらからパターンを抽出し、それらを記録したファイルを新たに 14 個作成した。また、投球の流れを分析するうえで、パターン長が長いパターンを抽出するため、支持度を 0.1 (10%) に統一した。SPMF でパターン抽出を行い新たに作成された出力データの一部の例をFig. 2 で示す。入力データは -1 を投球ごとに格納されているため、-1 が 2 つ表示されているパターンはパターン長が 2 であり、投球の流れを 2 球以上記録したパターンということになる。Fig. 2 のデータでは入力データ内の打席の中の 10% ではアウトコースのボール球が使われていること、その後にコース真ん中や低めのボール球が投げられていることなどが記録されている。

3. 3 抽出されたパターンの分析

最後に抽出されたパターンを分析した。投球の流れを研究することと、それらがアウトにとった打席とヒットを打たれた打席に違いがあるのかを調査するという点で、アウトにとった打席から抽出されたパターンとヒットを打たれた打席から抽出されたパターンで差集合をとり、さらにその中から、パターン長を 3 のものに絞って分析を行った。差集合やパターン長の限定は全てPythonで行い、アウトのパターン集合からヒットのパターン集合を引いたものと、ヒットのパターン集合からアウトのパターン集合を引いたものの 2 つのファイルを作成した。これにより、アウトをとった打席のみに現れたパターンとヒットを打たれた打席のみに現れたパターンを投手ごとに記録することができた。

アウトをとった打席のみに現れたパターンが記録され

```
straight middleh outs -1 fbrake middleh outs -1 -2
straight highb outb -1 straight highs outs -1 sbrake middleh middlec -
straight highb outb -1 sbrake lowb ins -1 straight middleh middlec -
fbrake lowb middlec -1 straight lows outs -1 fbrake lowb outb -1 sbr
straight middleh ins -1 straight highb inb -1 fbrake middleh outs -1
straight middleh inb -1 fbrake lowb ins -1 fbrake lowb middlec -1 st
sbrake middleh outs -1 fbrake highs ins -1 sbrake middleh outs -1 -2
fbrake lowb outs -1 straight highb middlec -1 fbrake lows middlec -1
fbrake middleh outs -1 -2
straight lows outs -1 straight lowb inb -1 fbrake lows outs -1 fbrak
sbrake middleh ins -1 -2
sbrake middleh outs -1 straight highs middlec -1 -2
fbrake lows outs -1 -2
fbrake lows ins -1 fbrake lows middlec -1 -2
```

Fig. 1. 収集されたデータの一部

```
straight middleh outs -1 fbrake middleh outs -1 -2
straight highb outb -1 straight highs outs -1 sbrake middleh middlec -
straight highb outb -1 sbrake lowb ins -1 straight middleh middlec -
fbrake lowb middlec -1 straight lows outs -1 fbrake lowb outb -1 sbr
straight middleh ins -1 straight highb inb -1 fbrake middleh outs -1
straight middleh inb -1 fbrake lowb ins -1 fbrake lowb middlec -1 st
sbrake middleh outs -1 fbrake highs ins -1 sbrake middleh outs -1 -2
fbrake lowb outs -1 straight highb middlec -1 fbrake lows middlec -1
fbrake middleh outs -1 -2
straight lows outs -1 straight lowb inb -1 fbrake lows outs -1 fbrak
sbrake middleh ins -1 -2
sbrake middleh outs -1 straight highs middlec -1 -2
fbrake lows outs -1 -2
```

Fig. 2. 抽出されたパターンを記録したデータの一部

```
straight -1 sbrake -1 lowb -1
middlec -1 fbrake -1 fbrake -1
straight -1 middleh -1 sbrake -1
outs -1 fbrake -1 middleh -1
middlec -1 straight -1 outs -1
fbrake -1 fbrake -1 outs -1
middlec -1 straight -1 lowb -1
middleh -1 fbrake -1 fbrake -1
```

Fig. 3. アウトにとった打席のみに現れたパターン長 3 のパターンの一部

```
sbrake -1 lowb -1 middlec -1
sbrake -1 sbrake -1 middlec -1
straight -1 sbrake -1 fbrake -1
lowb -1 fbrake -1 straight -1
fbrake -1 straight -1 middlec -1
middleh -1 straight -1 middleh -1
outs -1 sbrake -1 middleh -1
outs -1 lowb -1 middleh -1
outs -1 sbrake -1 middlec -1
```

Fig. 4. ヒットを打たれた打席のみに現れたパターン長 3 のパターンの一部

たデータの一部をFig. 3、ヒットを打たれた打席のみに現れたパターンが記録されたデータの一部をFig. 4 に示す。Fig.3 のデータの 1 行目は「straight -1 sbrake -1 lowb -1」と記録されており、ストレートの後に遅い変化球、その後に低めのボール球という投球パターンが記録されていることが分かる。Fig. 5 のデータの 1 行目は「sbrake -1 lowb -1 middlec -1」と記録されており、遅い変化球の後、低めボール球、その後にコース真ん中という投球パターンが記録されていることが分かる。これらのデータを用いて分析を行う。

次に分析手法について説明していく。今回はパターン自体の分析ではなくパターンを構成する各要素の傾向分

析を行った。パターン長3のパターンのうち最初に出てくる属性を1球目、二番目に出てくる属性を2球目、三番目に出てくる属性を3球目とし、各球でどのような属性が多いのかを分析していく。今回は各要素の属性危険度分析と球種分析を行った。

まずは属性危険度分析について説明していく。各属性の6投手全体の被打率の計算を行った。被打率の計算方法は6投手それぞれの各属性の被打率を投球数で重みづけするというものである。また、各投手の属性ごとの被打率は「データで楽しむプロ野球」[8]を参照した。6投手の平均被打率の計算式を式(1)に示す。

$$\text{平均被打率} = \frac{(\text{投球数} \times \text{各属性の被打率}) \text{の} 6 \text{ 投手分の和}}{\text{投球数の} 6 \text{ 投手分の和}} \quad (1)$$

式(1)で算出される平均被打率が高い3個の属性を危険、それ以外のストライクゾーンの属性と球種属性の6個の属性を安全、ボールゾーンの4個の属性をボール球に分類した。これら3種の危険度が1球目、2球目、3球目でどのように変化する傾向があるかを分析した。分析結果は棒グラフ形式で可視化できるようにし、投球の流れや変化を分かりやすく表示した。

次に球種分析について説明していく。各要素での球種属性の割合が投球を重ねるごとにどのように変化する傾向があるかを調査した。そのために、各要素での各球種属性の数が重要になるが、各要素での球種属性の数が一定ではないことから、各球種属性から球種属性全体の数を割ったものを算出した。これらの値を棒グラフで可視化し、アウトにとった打席とヒットを打たれた打席では

投球の球種の流れにどのような違いがあるのかを調査した。

4. 結果

4.1 各属性の危険度と抽出されたパターンの関係

各属性に先述した計算方法で6投手全体の被打率を計算し、その属性の危険度を分類することを行った。計算結果をTable.4で示す。「middleh」、「lows」、「middlec」が、被打率が高い上位3つの属性となった。よって、高さ真ん中、低めストライク、コース真ん中の3つの属性が、6投手がヒットを打たれる傾向が比較的高い属性であるということが分かった。これらの属性を危険に分類し、その他のストライクゾーンの属性と球種属性である「straight」、「fbrake」、「sbrake」、「highs」、「ins」、「outs」、つまり、ストレート、速い変化球、遅い変化球、高めストライク、インコースストライク、アウトコースストライクの6つの属性を安全に分類した。また、ボールゾーンの属性である「highb」、「lowb」、「inb」、「outb」、つまり高めボール球、低めボール球、インコースボール球、アウトコースボール球の4つの属性は被打率を計算するうえでデータが不足しており、確率が極端な値をとる可能性が高いため、全てボール球に分類した。各属性の危険度の定義表をTable.5に示す。また、表の危険度の配色は棒グラフの配色と対応しているものとする。

Table.5の定義表から棒グラフを作成し、投球の流れで危険度がどのように移り変わっていくのかを考察していく。まずは6投手全ての投球データから抽出されたパターンで各要素の危険度を分析した。アウトにとった打席から抽出されたパターンの棒グラフをFig.5、ヒットを打たれた打席から抽出されたパターンの棒グラフをFig.6に示す。

データ収集を行った6投手の全体の傾向としてアウトをとった打席では投球の危険度が投げていくにつれて低くなってきているということが分かり、ヒットを打たれた打席では1球目、2球目ではボール球、安全のように危険度の低い球を投げられているものの、続く3球目で危険度の高い球を投げてしまっているという傾向が出ることが分かった。アウトにとった打席では危険に分類された投球が投げるごとに少なくなっている分、ボール球に分類された投球が投げるごとに増えていた。投手カウント有利の時にはボール球を振らせることが有効であるという点でこの結果が現れたと考えられる。また、安全に分類された投球の割合はほぼ一定で全体の投球割合でも多くを占めていた。このことから、アウトにとった打

Table. 4. 各属性の被打率

属性	被打率
straight	0.258
sbrake	0.202
middleh	0.269
ins	0.242
outs	0.249
fbrake	0.228
highs	0.250
lows	0.261
middlec	0.290

Table. 5. 各属性の危険度の定義表

危険度	属性
ボール球	highb, lowb, inb, outb
安全	straight, fbrake, sbrake, highs, ins, outs
危険	middleh, lows, middle

席ではストライク先攻かつ危険度の低いゾーンを中心としたピッチングを展開していることが分かった。ヒットを打たれた打席では3球目の勝負球に危険に分類された投球の割合が高い。1球目、2球目の危険に分類された投球の割合はアウトをとった打席に比べて低いことが分かった。1球目と2球目に安全な球を投げていても、勝負球で危険なゾーンに投げてしまったときに打たれるケースがあることが推測される。また、1球目と2球目にボール球を投じている割合が、アウトをとった打席よりもヒットを打たれた打席に多く、ボールカウントが増え、ストライクゾーンでの勝負をせざるを得ない場合にそれを狙われて打たれるケースが存在すると推測される。

4. 2 球種による分析

先ほどの各属性の危険度による分析において、球種属性の被打率に大きな差が生まれなかったことから、全て安全に分類されていた。そのため、球種によるパターン分析も行った。6 投手全ての投球のアウトをとった打席の投球データから抽出されたパターンの球種推移の棒グラフをFig. 7、ヒットを打たれた打席から抽出されたパターンの球種推移の棒グラフをFig. 8 に示す。

アウトをとったパターンではストレート中心の組み立てから、速い変化球を使うことが多いということが分かった。一方ヒットを打たれたパターンでは最後にストレートを打たれるケースが多いことが分かった。アウトをとった打席ではストレートの割合が投球ごとに下がっていくことが明らかになった。特に1球目は投球のほとんどがストレートであり、ストレートで初球にストライクカウントを取ることが可能であれば、打者をアウトにとれる確率は上がると考えられる。一方、ストレートの割合が下がっていくのと同時に速い変化球の割合が上がっていた。ストレートの後に投じるフォークやカットボールなどの速い変化球は打者を惑わすうえで非常に効果的であると考えられる。一方ヒットを打たれた打席ではストレートを打たれているケースが多かったが、1球目と2球目の球種割合は遅い変化球、速い変化球、ストレートの3つともほぼ同じ値になった。ストレートは投手の投球割合の中で一番多いボールであることから、打者はそのボールを待っているケースが多く、1球目と2球目でどのような球種を投げて勝負球にストレートを選択することは危険であると推測される。アウトをとった打席でも勝負球をストレートにしているケースがみられることから、ストレートを勝負球にする場合はコースに対して精度よく投球することが重要と考えられる。

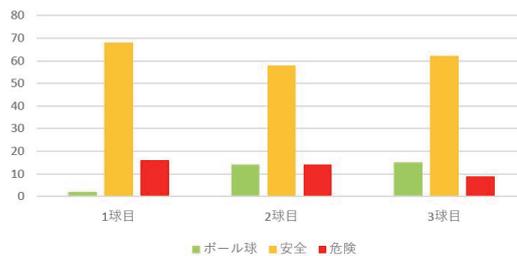


Fig. 5. アウトをとった打席から抽出されたパターンの危険度推移

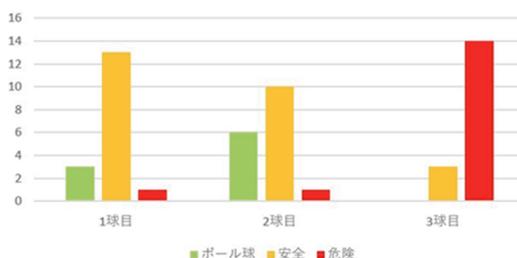


Fig. 6. ヒットを打たれた打席から抽出されたパターンの危険度推移

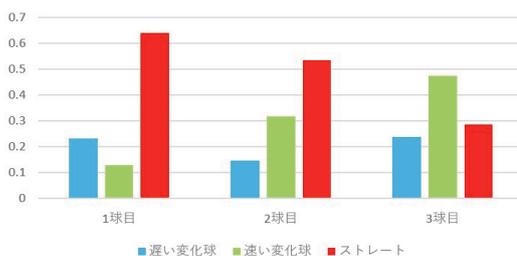


Fig. 7. アウトをとった打席から抽出された球種推移

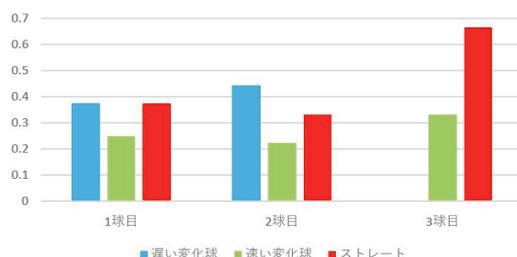


Fig. 8. ヒットを打たれた打席から抽出された球種推移

4. 3 イニング別のヒットパターン出現数

最後に抽出されたパターンがどのイニングで現れているかを調査した。SPMFの機能で各入力パターンをID化し抽出された出力パターンがどのIDで現れているかを表示するものがあり、それを用いてヒットパターンのイニング別の出現数を集計した。パリーグ6 投手全体の投球データから抽出されたヒットパターンのイニング別の出現数をFig.9 に示す。

一番ヒットパターンが現れた打席が多いイニングは 4

回であることが分かった。先発投手にとって4回は打者と2回目の対戦になることが多いため、投手のボールに打者が対応しやすいイニングであることからこのような結果が現れたと考えられる。また、7回以降、ヒットパターンの出現数が右肩下がりになっており、一番ヒットパターンが少ないイニングは9回であることが分かった。現代のプロ野球はリリーフの投手の能力も非常に高く、先発投手は5回から7回の間に降板することが多いため、登板数の少なさからこのような結果が出力されたと考えられる。また、7回のヒットパターンの出現数が比較的多いことから、7回でスタミナがなくなってしまうケースが推測され、投手交代のタイミングの1つの参考になり得ると考えられる。

5. まとめ

全体の傾向として投手が打者をアウトにとるためには、3つ目の要素である勝負球の精度を高めることが重要であると考えられる。「決め球に取って置きたいボール」という言葉がある通り、カウント球で最高のボールを投げても、勝負球に甘い球を投げるとヒットを打たれるケースが考えられる。また、アウトをとった打席では勝負球にボール球を使い、ヒットを打たれた打席では勝負球の前にボール球を使っていたことから、投手有利カウントで投球することが打者をアウトにとるうえで大切であることも確認できる。

球種分析では全体の傾向として、速い変化球を混ぜた攻めが有効であるということも分かった。メジャーリーグなどでも、ストレートと同じ軌道だが打者がスイングをかけてからは少し曲がる速い変化球が主流となりつつあり、それらを使った配球を用いることで打者をアウトにできる確率が上がると考えられる。また、1球目がほとんどの割合でストレートであることから、速い変化球の前に投じるストレートでストライクを取ることもアウトをとるうえで重要であると考えられる。また、ヒット

を打たれた打席での球種分析では全体の傾向として勝負球にストレートを投じていた。データ収集を行った投手のストレートの被打率が高い傾向にあったため、このような結果が現れたと考えられる。そのため、投手は勝負球でストレートを使う際には細心の注意を払う必要があると考えられる。

また、イニング別のヒットパターン出現数も調査し、全体的に2回、4回にヒットパターンが出現していた。今回はスタミナのある投手を中心にデータを収集したため、6回や7回にも多くのヒットパターンが現れた。これも、立ち上がり得意な投手と苦手な投手、スタミナのある投手とない投手によって結果が異なるため、投手の特徴を表す一つの指標になり得ると考えられる。

今後の展望として先述の通り、支持度を下げることやパターン長の調整を行うことでよりよい配球研究ができると考えられる。

参考文献

- [1] 堺田寛一郎, 川又泰介, 松田源立, “シーケンシャルパターンマイニングを用いた麻雀の捨て牌の傾向分析,” 情報処理学会第84回全国大会講演論文集2, pp.404-405, 2021.
- [2] 稲福和史, 伏見卓恭, 佐藤哲司, “最終球への配球推移に基づくキャッチャー成績分析,” 第17回日本データベース学会年次大会, 2019.
- [3] 小林裕哉, 大塚真吾, “プロ野球全球データから見る統計分析と実用性,” 第17回日本データベース学会年次大会, 2019.
- [4] 菊池啓太, 中島宣行, 綿田博人, “大学野球における配球について：カウント0-0における投球の分析,” 慶應義塾大学体育研究所紀要, vol.50, no.1, pp.41-51, 2010.
- [5] <https://baseball.yahoo.co.jp/npb/> (2022年5月24日アクセス)
- [6] <https://www.philippe-fourrier-viger.com/spmf/> (2022年9月15日アクセス)
- [7] 掌田津耶乃, つくってマスターPython：機械学習・Webアプリケーション・スクレイピング・文書処理ができる!, 技術評論社, 2019.
- [8] <https://baseballdata.jp/> (2022年11月3日アクセス)

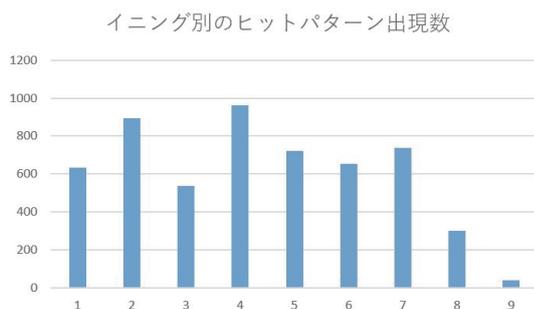


Fig. 9. イニング別のヒットパターン出現数