

# Jリーグスタッツデータと移籍データを用いた勝点の要因分析

2020SC007 藤田昇平 2020SC010 畠中響生

指導教員：河野浩之

## 1 はじめに

### 1.1 日本プロサッカーの現状

Jリーグの入場者数はJリーグ発足の1999年が361万人だったのに対し、2019年では1040万人にまで増加しており、日本においてサッカーは人気スポーツとなっている。[1] さらに最近では、日本でのサッカー人気の向上に伴い、サッカー界のスター選手のJリーグへの移籍が増加し、国外からも実績と人気のある選手がJリーグチームに加入している。実績や人気のある選手の移籍のみではなく、Jリーグチーム同士での移籍も活発に行われるようになり、2014年以降移籍数は増加傾向にある。2014年以降の平均移籍人数の推移を示したグラフを図1に示す。

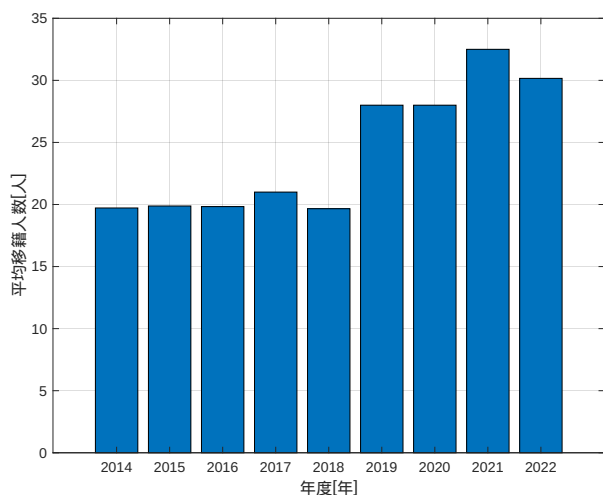


図1 1チームあたりの年間平均移籍人数の推移

図1はサッカー情報サイト「ゲキサカ」[2]からJ1リーグに所属するチームの移籍数を集計し、それぞれの年の1チームあたりの平均移籍数をまとめたものである。このグラフから2014年から2022年にかけて移籍人数は増加傾向にあることがわかる。

移籍数の増加に加えて、Jリーグにおいては、チーム人件費と成績には強い相関が確認された。経済産業省の資料[1]より、チームの人件費を増加することでチームの成績の向上につながるがわかっている。このチーム人件費に関連して、本研究ではJリーグにおける移籍とチーム成績の関係性に着目した。

## 2 サッカーデータ分析に関する先行研究

### 2.1 重回帰分析を用いた研究

吉田ら[3]は、サッカーにおいて勝敗に最も関係のある得点がどういった要因から生じるのか、スタッツデータと

の関連性を検討するためにJリーグのスタッツデータを用いてオープンプレー得点と得点について重回帰分析を用いて検討した。研究には、2019～2021年までのJ1、J2、J3のリーグ戦を研究対象として研究を行った。

オープンプレー得点の推移を調べるために重回帰分析を行った。ステップワイズ法を用いて変数選択を行い、オープンプレー得点を目的変数とし、スタッツデータのうち、攻撃回数、シュート数、枠内シュート数のような得点に関連のあるスタッツを説明変数として分析を行った。結果として有効な回帰式が得られ、その寄与率は38%であり、推定値の標準誤差は、約0.2点という結果を得た。このことから、枠内シュート数とペナルティエリアへの進入回数から、オープンプレー得点の38%が説明できるという結果を得た。

また、オープンプレー得点と同様に得点を目的変数とした重回帰分析を行った。結果として、寄与率47.9%、推定値の標準誤差、約0.2点となり、得点においても枠内シュートおよびペナルティエリアへの進入回数を増やすことで得点の期待値が高まると言える。

最終的に、オープンプレー得点と得点に関して、ペナルティエリアへの進入回数と枠内シュートを変数とした予測式が求められ、ペナルティエリアへの侵入回数11回、枠内シュート4回で1得点が期待できると示された。このことから、Jリーグにおける得点は枠内シュート数及び、ペナルティエリアへの侵入回数で予測できることが分かった。また、予測式による勝敗シミュレーションの精度としては1/3程度であったが、引き分けと判定した試合にはホームアドバンテージがあるのではないかと予想された。吉田らの研究においては、Jリーグ全体での分析であり、個々のチームの特色が反映されておらず、チームごとでの戦術による影響についての検討がなされていなかった。

### 2.2 探索的因子分析を用いた研究

徐ら[4]は、Jリーグにおいてパフォーマンスに関するビッグデータが、サッカーに求められる技術を反映していることを検証するために、その因子構造を明らかにすることを目的とした研究を行った。この研究では徐らが先行研究として上げた研究で用いられた45項目のスタッツデータを用いた。また、最も妥当であると考えられる因子構造を見つけるために、因子抽出法、回転法、因子数のハイパーパラメータにおいてすべての組み合わせを実行するプログラムを開発し、検証を行った。複数指定されたハイパーパラメータのすべての組み合わせで探索的因子分析を行うことで、単純構造指数をもとに有意と考えられる結果を優先して出力する関数の開発と実装(PAHFAの開発、実装)を

行った。15 回目の PAHFA 実装の結果、第 6 因子まで因子負荷量の基準と定めた 0.52 を満たした。さらに、累積寄与率は 79 % となり、結果として 14 変数、6 因子で構成されるという結果を得た。徐らの研究においては、ゲーム・パフォーマンス・ビッグデータは、ボール保持者の周囲の選手の動きはデータ化されておらず、ボール保持者の範囲内で解釈する必要がある。

### 2.3 サッカーに関するデータ分析の先行研究を踏まえて

これらの先行研究を踏まえて、先行研究である吉田ら [3] の研究は攻撃時の得点パターンの分析であり、チーム全体のパフォーマンスではない。また、徐ら [4] らの研究においても攻撃時のスタッツデータに関連した研究であった。これら先行研究は攻撃時に焦点を当てたデータ分析であり、チーム全体の成績を考慮して行われた研究ではなかった。また、サッカーのデータ分析に関する研究は 1 試合でのデータを対象としたものがほとんどであった。そこで本研究では、スタッツデータの利用という点と先行研究で用いられている重回帰分析の手法を参考に J リーグのスタッツデータ及びチームデータを用いて選手の移籍と勝点の関連性について研究を行う。本研究では J リーグのリーグ戦における勝点に焦点を当て、勝点に関連のあるスタッツがどのようなものかを分析し、勝点に深く関わるスタッツを導き出すことを目標とする。加えて、チームの成績が選手の移籍によってどのように変化するかを分析し、移籍がチームの成績に及ぼす影響の分析を行う。分析に勝点を用いた理由としては、勝点は年間を通してのチーム全体の成績を示すものであり、攻撃的な選手だけではなく、ゴールキーパーやディフェンダーなどの守備的なポジションを担う選手の移籍によるチーム成績の推移を評価できるからである。そのため、移籍と勝点に着目した研究を行う。

### 3 重回帰分析を用いた J リーグのスタッツ、移籍データに関する分析の提案

本研究では、サッカースタッツデータを提供しているサイト "footystats" にあるスタッツデータから説明変数の候補を挙げる。今回は選手の移籍とチームの勝点の関係について研究を行うため、目的変数を勝点として定めた。説明変数の候補、目的変数をまとめたものが表 1 である。

表 1 説明変数と目的関数

目的変数	勝点
説明変数	コーナーキック数 カード数 平均ポゼッション 枠内シュート数 ファウル数 前半リード数 先制点

説明変数の候補の選定にあたって明らかに勝点とは関係

ないと考えられるもの、他の説明変数と多重共線性が明らかであるスタッツは事前に候補から排除した。

上記で上げた説明変数の候補と 2014 から 2022 年度の間 J1 に所属していたクラブチームの勝点と各チームのスタッツとの相関を求める。ここでは 2022 年度の J1 リーグに所属する各チームの勝点を抜粋し、まとめたものを表 2 に示す。

表 2 2022 年度 J1 所属クラブの勝点

チーム名	J1 各チームの勝点
京都サンガ F.C.	36
清水エスパルス	33
コンサドーレ札幌	45
セレッソ大阪	51
ジュビロ磐田	30
アビスパ福岡	38
湘南ベルマーレ	41
名古屋グランパス	46
横浜 F・マリノス	68
浦和レッズ	45
サガン鳥栖	42
ガンバ大阪	37
鹿島アントラーズ	52
サンフレッチェ広島	55
ヴィッセル神戸	40
FC 東京	49
柏レイソル	47
川崎フロンターレ	66

表 2 に示した各チームの勝点と表 1 で設定したスタッツのデータとの相関 (2021 年 2022 年) を表 3 として示す。

表 3 勝点との相関

スタッツ	相関
コーナーキック数	0.45
カード数	-0.16
平均ポゼッション	0.54
枠内シュート数	0.61
ファウル数	-0.14
前半リード数	0.72
先制点	0.79

表 3 より、勝点との相関が高いスタッツとして枠内シュート数、先制点が高いとわかる。一方で、ファウル数、カード数と言ったような攻守どちらの場面においても発生するスタッツは勝点との相関が弱いことがわかる。今回は研究に用いるスタッツを表 3 であげたうち相関が 0.45 以上であるものを対象として研究を行う。表 4 として本研究に用いるスタッツを示す。

表 4 勝点との相関その 2

スタッツ	相関
コーナーキック数	0.45
平均ポゼッション	0.54
枠内シュート数	0.61
前半リード数	0.72
先制点	0.79

以降では、上記表 4 であげた 5 つのスタッツと "jleague.jp" より集めた各チームの移籍情報をもとに回帰分析を行う。

## 4 スタッツデータおよび移籍データに対する回帰分析

### 4.1 分析準備

本研究ではスタッツデータの取得先として "footystats" を利用し、データの加工、整理ツールとして Excel、データ分析ツールとして Exploratory を使用した。また、移籍人数のデータは "jleague.jp" で公開されている情報を元に自ら作成した。使用したサイト、ツール等をまとめたものを表 5 に示す。

表 5 使用したサイト、ツール等

サイト名、ツール名	概要
footystats	Jリーグスタッツデータ収録先
Exploratory	データを探索する分析ツールであり、操作しやすく分析の種類、幅が広い
excel	データの加工、整理

本研究では主に 3 つの分析を行った。1 つ目の分析は、J1 リーグの 2014 ~ 2022 年分のスタッツデータを対象として、勝点とスタッツの影響度分析を行った。この分析によって勝点と関係の強いスタッツを結果として得られる。この分析結果、回帰式をもとに移籍の影響度を 2 つ目、3 つ目の分析では比較する。コーナーキック数は分析の過程で有意でなかったため説明変数から排除した。2 つ目の分析は、1 つ目の分析に総移籍人数のデータを含めた分析を行った。2021、2022 年の 2 年分のスタッツデータと移籍データを対象として勝点とスタッツ、総移籍人数の影響度分析を行った。また総移籍人数と勝点の単回帰分析についても行った。この分析は、そもそも移籍人数と勝点との影響度の関連が存在するのか確かめるために行った。ここでの総移籍人数とは移籍による加入数と放出数の合計人数である。3 つ目は、勝点と実績のある選手の移籍との回帰分析を行った。2 つ目の分析と同様にして、1 つ目の分析に加えて実績ある選手の移籍データを含めた重回帰分析、勝点と実績ある選手の移籍データにおいて単回帰分析を行った。実績ある選

手の移籍データは、過去に J1 リーグで 1 シーズンにおいて半数以上の出場記録のある選手、獲得した外国人選手をカウントしたデータである。

### 4.2 勝点に対する影響度分析

スタッツデータに関しても行った分析結果を示す。

表 6 スタッツデータのための重回帰分析の結果

変数	係数	標準誤差	t 値	P 値
切片	-18.46	6.787	-2.7	0.007
先制点	1.60	0.248	6.4	0.000
前半リード数	0.77	0.285	2.7	0.008
平均ポゼッション	0.41	0.160	2.6	0.011
枠内シュート数	0.075	0.029	2.6	0.011

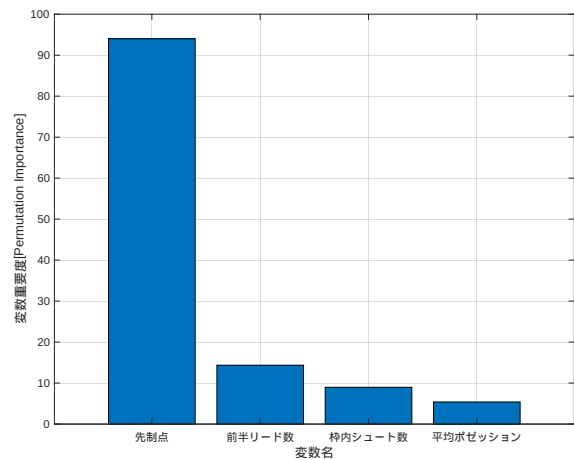


図 2 スタッツデータに対する重回帰分析

1 つ目の実験の結果として、図 2 から先制点を取ることが特に重要であるということがわかる。また、表 6 より、t 値が 2 より大きく、p 値が 0.05 より小さいという有意水準を全ての説明変数が満たしていることがわかる。

### 4.3 移籍を含めた分析結果

2 つ目の実験として、総移籍人数のデータを分析に含めた分析結果を表 7、表 8 に示す。

表 7 移籍総数を含めた重回帰分析の結果

変数	係数	標準誤差	t 値	P 値
切片	11.24	3.297	3.4	0.000
先制点	1.92	0.261	7.4	0.000
前半リード数	0.98	0.308	3.2	0.001
移籍総数	-0.15	0.080	-1.8	0.068

表 7 と表 8 から移籍総数は有意な説明変数ではなく、単に選手の入れ替わりを多くすることが勝点につながるわけではないことがわかる。

3 つ目の実験として、移籍した年から遡り、過去 5 年以内に同リーグ (J1) 内で年間で過半数の試合出場記録のある

表 8 移籍総数単回帰分析

変数	係数	標準誤差	t 値	P 値
切片	51.969	3.374	15.4	3.013
移籍総数	-0.18	0.135	-1.3	0.178

選手を実績のある選手とし、その選手を対象に重回帰分析と単回帰分析を行った。結果を表 9、表 10 として示す。

表 9 実績ある選手を対象に加えた重回帰分析

変数	係数	標準誤差	t 値	P 値
切片	13.99	6.67	2.10	0.044
加入 (実績あり)	-0.93	0.542	-1.7	0.094
先制点	2.16	0.506	4.3	0.000
前半リード	0.60	0.611	1.0	0.336

表 10 実績ある選手を対象に加えた単回帰分析

変数	係数	標準誤差	t 値	P 値
切片	64.34	5.95	10.83	0.000
移籍 (実績あり)	-2.44	0.89	-2.76	0.009

上記、表 9、表 10 は実績のある選手を説明変数に加え、重回帰分析を行った結果である。表 10 単回帰分析において有意水準を満たしていることがわかる。

2 つ目、3 つ目の分析は、移籍データが 2 年分のみであるため、分析に利用できるデータ数を考慮し、重回帰分析に用いる説明変数は実績のある選手の移籍データ、先制点、前半リード数の 3 つで分析を行った。

補足として上記の分析結果は、全て多重共線性の指標である VIF 値が 10 以下で多重共線性はなかった。また、変数重要度は Permutation Importance というモデルを用いて求めた。

## 5 回帰式からの予測と考察

勝点を構成する要因として先制点、ポゼッション、枠内シュート数、前半リードが勝点到深く関わっていることが確かめられた。回帰式の係数に着目すると、先制点を取る試合を 1 試合行うだけで勝点を約 1.6 獲得できる事がわかる。このことから、先制点を取ることが特に重要だと考えられる。1 つ目のスタッツデータに関する重回帰分析の結果を 2014 年度の J1 所属クラブから 3 チーム抜粋し、今回求めた回帰式にデータを代入すると表 11 のようになる。

表 11 回帰式への代入結果

チーム名	2014 年の順位	実際の勝点	予測結果
徳島	18 位	14	18.5
名古屋	10 位	48	46.9
浦和	2 位	62	67

表 11 は、得られた回帰式にデータを代入したものと実際の 2014 年度のリーグ戦の結果を比較した表である。回帰式への代入は上位、中位、下位のチームを抜粋して行った。代入結果は、勝点の誤差 5 点以内という結果を得た。また決定係数も約 0.68 で勝点を予測する有意な回帰式を得た。移籍を含めた分析結果からは、単に多くの選手を移籍によって入れ替えることでは勝点到直結しないことがわかった。さらに、本研究で行った実績のある選手の加入と勝点の単回帰分析においては、有意な回帰式が得られ、実績のある選手が一人加入するごとに勝点が約 2.4 減少するという結果になった。相関係数においても唯一、-0.4 以上と他の移籍データの勝点との相関係数と比べて高い結果であった。ただし切片が約 64(勝点)であることから、上位チームでの想定が妥当だ。また決定係数は約 0.17 であり、慎重に解釈する必要がある。この結果から実績のある選手の加入がチームのスタイルや戦術に大きな影響を与え、チームとしては勝点を積み上げられない、つまりチームとしては弱くなってしまうことが示唆される。

## 6 結び

J1 リーグのスタッツデータの分析の結果、勝点を構成する要因として先制点、ポゼッション、枠内シュート数、前半リードが勝点到深く関わっていることが確かめられた。移籍に関する分析の結果、実績のある選手の加入がチームのスタイルや戦術に大きな影響を与え、チームとしては勝点を積み上げられない、チームとしては弱くなってしまうということが示唆される。本研究で目標としていた移籍の提案としては、実績のある選手の獲得ではなく、実績を考えずに、前半から攻撃的なサッカーを体現できる選手を獲得することが重要である。

## 参考文献

- [1] 経済産業省, “J リーグの成長戦略について,” <https://www.meti.go.jp/shingikai/> (参照 2023. 10. 22)
- [2] “ゲキサカ,” <https://web.gekisaka.jp/> (参照 2023. 10. 25)
- [3] 吉田充, “サッカーのスタッツデータを対象とした得点要因の推定,” 北海学園大学経営論集, pp.1-13, 2022.
- [4] 徐広孝・中西健一郎・青木優, “J リーグの攻撃プレーにおけるゲーム・パフォーマンス・ビッグデータの因子構造-探索的因子分析における実用的プログラムを適用して-,” 静岡産業大学論集, Feb. 2022, pp.65-84.
- [5] j リーグ公式サイト, “2022 シーズン移籍情報,” <https://www.jleague.jp/special/transfer/2022/>, (参照 2023. 8. 31)
- [6] サッカーデータとスタッツの統計分析サイト-footystats, “J リーグスタッツデータ,” <https://footystats.org/jp/>, (参照 2023. 8. 11)