

J リーグ FW 選手の統計的分析

2012SE290 谷田川涼

指導教員：木村美善

1 はじめに

サッカーというスポーツは 11 人で行い、その中で大きく分けて GK, DF, MF, FW というポジションが存在する。相手より多く点をとる事が勝敗を決める要因であり、得点をとる役割の多くは、FW が担っている。各選手の役割が成熟したチームになるほど明確化されており、監督の考えやチーム全体での理想の試合の進め方がある中、FW の選手の役割や各選手の強み、また勝つために何が必要かを卒業研究のテーマとした。

2 データについて。

分析対象の選手データは 2014 年度 J リーグでの各選手の結果である。選手の制約条件としては、全 34 試合の 3 分の 2 以上出場し選手登録が FW の選手とした。

CB (Chance Building) とは、「選手 (またはチーム) が試合を通じてどれだけチャンス機会を構築したか」を独自のロジックにより数値化した指標である。上記のチャンス機会とは、今回は「シュート機会への貢献」という観点での評価とする。考え方は

$$A \times B = \text{評価} \quad (1)$$

とし A は (該当プレー) \times (エリア別シュート) 達成率, B は該当エリアにおけるプレーの難易度評価とする。B の該当エリアにおけるプレー難易度はプレー別リーグ平均成功率のデータを使い、範囲と母数は 2008 年~2014 年の J1, J2, J3 の各リーグ戦のみのデータである。

3 主成分分析

2014 年度の以下の成績を用いて選手の特徴を判別するために主成分分析を行った。出場回数, 出場時間, ゴール, アシスト, 身長, パス CB, ドリブル CB, 攻撃 CB, シュート回数, 守備 CB, 敵陣空中戦, 自陣空中戦, シュート成功率, ボール奪取力, パスチャンス力, ドリブルチャンス力, クロスチャンス力, 守備力の 18 個を変数として用いた。第 5 主成分までで累積寄与率が 80 % を超えているため、第 5 主成分までを用いて分析した。

・第 1 主成分 (寄与率 31 %)

正：身長, 守備, 敵陣, 自陣空中戦

負：攻撃, パス, ドリブル

第 1 主成分は、守備的な FW が攻撃的な FW を分ける形となった。プレーするエリアは、プレー場面によって変動し、敵陣の深いエリア, 中央, 自陣の深いエリアとなる。

・第 2 主成分 (寄与率 18 %)

正：ゴール, ドリブル, シュート回数, パスチャンス力, ドリブル CB

負：守備, ボール奪取, 自陣空中戦

第 2 主成分は、ドリブルをプレーの軸とし、敵陣深くの危険なエリアで攻撃主体のプレーをメインに行うかどうかを表す。

・第 3 主成分 (寄与率 18 %)

正：敵陣空中戦, 身長, ゴール

負：ドリブル, パス, ボール奪取, 守備

第 3 主成分は、天性的なフィジカルを活かした FW かどうかを特徴づけるものとなった。敵陣深くに陣取り、カウンターやセットプレーのハイボールを狙うプレーをするかどうかを表す。

・第 4 主成分 (寄与率 8 %)

正：ドリブル, シュート成功率, アシスト

負：出場回数, 出場時間, ボール奪取

第 4 主成分は、献身的に動き数字に残らないプレーにも関わる FW を分ける形となった。エリアを広く使い短距離, 長距離を走り、チームに献身的なプレーをする選手かどうかを表す。

・第 5 主成分 (寄与率 7 %)

正：パス以外の攻撃に関わる変数

負：パス, シュート成功率

第 5 主成分はドリブルや空中戦を行い、前を向けばシュートを打つ様な、良い意味でエゴイストな FW を分ける形となった、エリアは幅広く敵陣を使うと考える。

3.1 考察

主成分分析から、選手が攻撃的か守備的かの区別や主としてプレーする場所がどこかを分析することが出来た。身長が高い選手は攻撃的な選手の中でも特に攻撃的な選手と予想していたが、セットプレーの度に守備をすることで自陣空中戦と守備力が伸びて守備的な選手と分類された選手もいた。それは同じような選手でも所属チームの戦術や状況的により守備をする選手としない選手に分かれるからだと思う。

4 チームの選出

分析を行うチームを選択するにあたっては得点や失点だけでなく得点パターン, 失点パターンなど 43 個の説明変数を用いてクラスター分析を行った。地元チームである名古屋グランパスエイト, 距離が近く似たチームと分類された優勝チームのガンバ大阪と降格した大宮アルディージャの 3 つのチームを選出した。

5 重回帰分析 (ガンバ大阪)

全 34 試合の各試合の結果から目的変数を勝敗とし、得点数 (x_2), 会場 (x_3), 天候 (x_4), 試合間隔 (x_5), シュート数

(x_6), 交代人数 (x_7), 支配率 (x_8), オフサイド (x_9), 30m ラインの進入回数 (x_{10}), 先取点 (x_{11}), パス回数 (x_{12}), FW が点をいれたか (x_{13}), を説明変数とした。勝敗, 会場, 天候, 試合間隔, 先取点, FW が点をいれたかななどの説明変数は, ダミー変数とし, 選手の分析の際に用いたデータとは重複していない。

5.1 回帰式

AIC に基づく変数選択を行ない表 1 の結果を得た。

表 1 重回帰分析結果 (ガンバ大阪)

	回帰係数	t 値	p 値
切片	0.0824	0.126	0.9006
x_2	0.1613	3.939	0.0005
x_7	- 0.2449	- 1.791	0.8419
x_8	0.0303	1.856	0.0740
x_{11}	0.4440	3.330	0.0024
x_{12}	- 0.0017	- 1.668	0.1065

決定係数は 0.727, 自由度調節決定係数は 0.6783。ガンバ大阪は攻撃的なチームであり, ドリブルやロングボールなどを活かした攻め方が理想的であることが支配率とパス数の正負の関係性から読み取ることが出来た。

5.2 先取点と勝敗の関係性

2014 年のリーグ戦全 34 試合の中, 0-0 の点が動かなかった試合がガンバ大阪は 2 試合存在したため, その試合を除き, 全 32 試合の試合データを基に, 先に点を獲得したのか, されたのか, またどのポジションの選手が点を取ったのか, 取られたのか, その時間帯などからガンバ大阪の勝敗を分析する。目的変数を勝敗とし, 全 90 分の時間帯を四等分する。なお, ロスタイムは 90 分とする。得点者のポジションも FW, MF, DF と分類する。得点パターンと失点パターンを考えるため, 説明変数は全 14 個ですべてダミー変数となる。AIC に基づく変数選択を行い, 表 2 の結果を得た。

表 2 先取点と勝敗 (ガンバ大阪)

	回帰係数	t 値	p 値
切片	0.0909	0.858	0.3984
x_2	0.7091	3.741	0.0009
x_3	0.7841	4.802	5.2×10^{-5}
x_4	0.6591	3.212	0.0034
x_5	0.9091	4.431	0.0001

決定係数は 0.5681, 自由度調節決定係数は 0.5041 である。ガンバ大阪は, この年守備的面も強くリーグでは 2 番目に失点が少ない, また先取点も遅い時間だと逃げ切りやすく, 早い時間でも追加点が狙える, 優勝チームらしい欠点のないチームといえる結果だ。

6 ロジスティック回帰分析 (ガンバ大阪)

確率のように範囲 $[0, 1]$ の値をとる目的変数を説明変数群 x_1, x_2, \dots, x_p で説明をするとき, 通常重回帰分析では説明変数群の合成変数のとる値が $[0, 1]$ の外に出てしま

い適切ではない。なので, 下式のように説明変数群の合成変数をロジスティック関数にすることで, その値域の範囲を $[0, 1]$ に収めることができる。

$$y = \frac{\exp(b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p)}{1 + \exp(b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p)}$$

2014 年のリーグ全 34 試合でガンバ大阪が先取点を獲得した試合が 20 試合存在する。先取点を獲得した時間帯 (x) に対して勝つ確率をロジスティック回帰した結果が表 3 である。

表 3 ロジスティック回帰結果 (ガンバ大阪)

	回帰係数	t 値	p 値
切片	0.6016	0.515	0.606
x_1	0.0320	0.983	0.326

ロジスティック曲線では, 1 分頃から上がり始め, 10 分頃には勝つ確率は 1 の値に達している。それは, 先取点を獲得した 20 試合中負けた試合が 3 試合のみということからだと考えられる。重回帰分析の考察の通り, どの時間帯でゴールをきめても逃げ切れる, また追加点を決められることができるチームであることから試合開始 1 分に点を獲得しても 50% の勝率があるといえる。

7 おわりに

本研究を終えて, サッカーというスポーツは 1 点の重みが多い。優勝チームは, 先取点を決めていた試合が多く, 負けにくいということであると思った。順位が真ん中のチームやデータを集める試合数を監督が変わる間などもっとデータ数を増やすことや他の分析方法を用いることでより明確に分析ができ, 先取点と勝敗の関係性に違いが現れるのではないかと思う。本研究のデータを集める際に利用した WEB サイトのようにサッカーに統計学が用いられている。最近スポーツ中継をみるとサッカーだけではなく野球など様々なスポーツに統計学が用いられるようになってきている。スポーツ統計学を理解し, スポーツ経験者の観点や統計学の観点からもスポーツを楽しめるようになれば今後の人生がより有意義になるものと考えられる。

参考文献

- [1] 井沢匡輝: OPTA による J リーガーの統計的分析, 南山数理情報学部情報システム数理学科卒業論文要旨集, 2002.
- [2] 週刊サッカーダイジェスト増刊 2014 J リーグ総集編, 日本スポーツ企画出版社, 2014
- [3] サッカーをデータで楽しむ—Football LAB
<http://www.football-lab.jp/>
- [4] 中村永友: 多次元データ解析法, 共立出版, 2009