

【卒業論文概要】

サッカーチームが勝利するための統計分析
～2014年度から2016年度までのJ1リーグを対象として～

14H3010 内海政広

1. はじめに

現代サッカーでは、あらゆるデータが得られる。データを基にして、チームと監督は戦略を組むのである。そして、どのような戦略をとっても求められるものは得点である。

そして、サッカーを統計的に分析されている研究の多くは、平均を求めたり試合の中で発生したプレーの割合を求めたりするなどの記述統計によるものであり、得られるデータから推測を行う推測統計があまりなされていないのが現状である。

試合に勝利することが必要とされ、その中でも試合で得点することの重要性が以前までよりも増していると思われる。したがって、得点に対して推測統計を用いて、分析を行う必要性があると考えられる。

本研究では、サッカーのJリーグによって公式に公開されているデータから統計分析をすることで、得点を決めるための要因を検討することを目的とする。

2. モデルとデータ

1節では、分析するモデルについて詳しく説明する。2節では分析をするために取り扱うデータについて説明する。

2.1 分析モデル

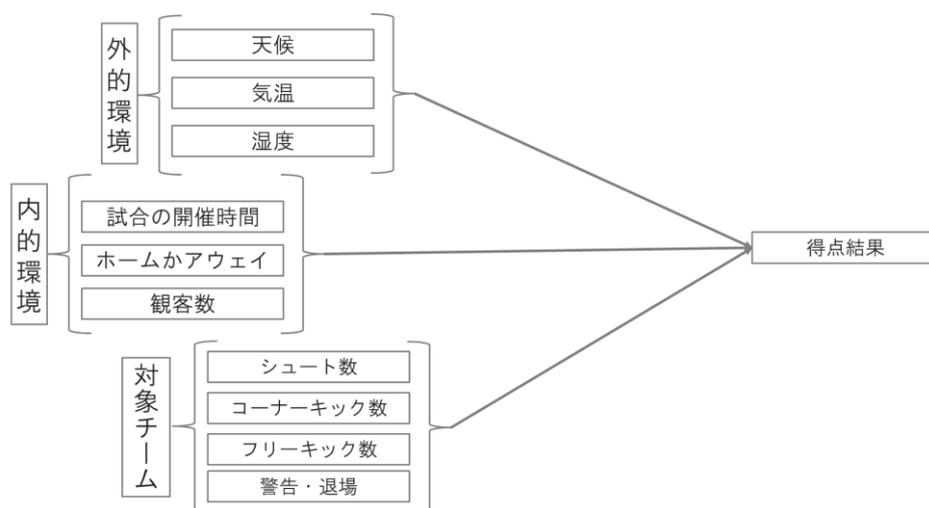


図1 本研究における分析用データ

図1に示すように、(i)外的環境とは、天候、気温、湿度のことを指す要因であると定義する。(ii)内的環境とは、試合の開催時間、ホームでの試合かアウェイでの試合かといった状

況や観客数のように試合が行われる状況を表すものであると定義する。(iii)試合データとは、総シュート数、セットプレー数（コーナーキック数、フリーキック数）、警告数、退場者数などの各選手に依存する要因であると定義する。

本研究では、チームの得点はポアソン分布に従うと仮定する。このとき、チーム i が対戦相手 j と試合したときの得点 Y が y_{ij} となる確率 P は式(1)のように書くことができる。

$$P[Y = y_{ij} | X] = \frac{\exp(-\lambda_{ij}) \cdot \lambda_{ij}^{y_{ij}}}{y_{ij}!} \quad \dots (1)$$

ここで、 Y は得点結果に関する確率変数であり、 y_{ij} はその実現値であり、 $y = 0, 1, 2, \dots$ と正の整数の値をとる。 λ_{ij} は得点の期待値であり、 X は説明変数のベクトルである。 X には、前述の外的環境、内的環境、その他の試合データに関する変数が相当する。

分析対象のチームを **team** として、対戦相手のチームを **opponent** とする。**team** と **opponent** の各要素はカテゴリカルデータであり、ダミー変数として定義する。**team** の各要素からなるベクトルを **team** とし、**opponent** の各要素からなるベクトルを **opp** とすると、**team**、**opp** は次のように書くことができる。

$$\begin{aligned} \mathbf{team} &= (\delta_1^t, \delta_2^t, \dots, \delta_i^t, \dots, \delta_{14}^t) \\ \mathbf{opp} &= (\delta_1^o, \delta_2^o, \dots, \delta_j^o, \dots, \delta_{14}^o) \end{aligned}$$

ここで、 δ_i^t は分析対象チームが i のときに 1 の値をとり、その他のときに 0 の値をとるダミー変数とし、 δ_j^o は対戦相手のチームが j のときに 1 の値をとり、その他のときに 0 の値をとるダミー変数とする。

得点を説明する要因からなるベクトル \mathbf{z} を以下のように定義する。

$$\mathbf{z} = (\mathbf{z}^o, \mathbf{z}^t, \mathbf{z}^G)$$

ここで、 \mathbf{z}^o は外的環境を表す変数ベクトルであり、 \mathbf{z}^t は内的環境を表す変数ベクトルであり、 \mathbf{z}^G はその他の試合データに関する変数ベクトルである。

そして、 $\boldsymbol{\alpha}$ を **team** の係数ベクトル、 $\boldsymbol{\beta}$ を **opp** の係数ベクトル、 $\boldsymbol{\gamma}$ をそのほかの変数の係数ベクトルとして以下のように定義する。

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_i \\ \vdots \\ \alpha_{14} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_{14} \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} \gamma^o \\ \gamma^t \\ \gamma^G \end{pmatrix}$$

ここで、 $\boldsymbol{\gamma}^o$ は外的環境を表す変数それぞれの係数からなるベクトルとする。同様に、 $\boldsymbol{\gamma}^t$ は内的環境を表す変数それぞれの係数からなるベクトル、 $\boldsymbol{\gamma}^G$ は試合データに関する変数それぞれの係数からなるベクトルである。

このとき、得点の期待値 λ_{ij} は *intercept* を定数項として、次の式で特定化する。

$$\lambda_{ij} = \exp(\text{intercept} + \mathbf{z} \cdot \boldsymbol{\gamma} + \mathbf{team} \cdot \boldsymbol{\alpha} + \mathbf{opp} \cdot \boldsymbol{\beta}) \quad \dots (2)$$

3章では、式(2)の推定を行い、外的環境や内的環境、その他の要因が得点に与える影響について考察される。併せて補助的な分析として、各種環境要因のうち、外的環境のみの効果を強調する場合、内的環境のみの効果を強調する場合、試合データのみの効果を強調する場合の3種類のモデルについても検討するが、それぞれのモデルは以下のように特定化する。

すなわち、外的環境のみの効果を強調する場合の推定式は以下のように記述される。

$$\lambda_{ij} = \exp(\text{intercept} + \gamma_1^0 \text{weather} + \gamma_2^0 \text{temperature} + \gamma_3^0 \text{humidity} + \alpha_i + \beta_j) \quad (5)$$

同様に、内的環境のみの効果を強調するモデルは次の式のように特定化する。

$$\lambda_{ij} = \exp(\text{intercept} + \gamma_1^1 \text{HorA} + \gamma_2^1 \text{audience} + \gamma_3^1 \text{HorA} \cdot \text{audience} + \alpha_i + \beta_j) \quad (6)$$

最後に、試合データのみの効果を強調するモデルは次の通りである。

$$\lambda_{ij} = \exp(\text{intercept} + \gamma_1^G \text{teamyellow} + \gamma_2^G \text{oppyellow} + \gamma_3^G \text{teamred} + \gamma_4^G \text{oppred} + \gamma_5^G \text{teamshoot} + \gamma_6^G \text{oppshoot} + \gamma_7^G \text{teamck} + \gamma_8^G \text{oppck} + \gamma_9^G \text{teamallfk} + \gamma_{10}^G \text{oppallfk} + \alpha_i + \beta_j) \quad (7)$$

2.2 データ

データの入手に関して本分析に利用するデータは、Jリーグの公式サイト「成績・データ」にある「日程・結果」より収集した。2014年から2016年までの複数年分のJリーグにおけるデータを収集する。このとき、2014年度は1シーズン制を採用していた。しかし、2015年・2016年度ではファーストステージ、セカンドステージからなる2シーズン制を採用していた。この2年間においても、年間順位は最終的に優勝が決まるチャンピオンシップにおいても優遇措置がなされていたため、分析に対して影響はないとしてデータの収集を行った。

3.推定結果

外的環境、内的環境、試合データを加味した上での推定結果を表2に示す。

統計的に有意である変数は気温、分析対象チームが1人目の退場者を出した場合、分析対象チームの総シュート数と総コーナーキック数、対戦相手チームの総コーナーキック数、総フリーキック数の6変数である。

AICは3079.9という結果が得られた。疑似決定係数は0.192という結果が得られた。

表 2 外的環境、内的環境、試合データを加味した上での推定結果

		係数	標準誤差	z値	Pr(> z)
	定数項	-0.700	1.233	-0.568	0.570
対象 チーム	浦和レッズ	0.105	0.136	0.768	0.442
	鹿島アントラーズ	-0.116	0.136	-0.856	0.392
	柏レイソル	-0.024	0.144	-0.167	0.868
	サガン鳥栖	-0.220	0.154	-1.427	0.154
	川崎フロンターレ	0.012	0.131	0.089	0.929
	横浜F・マリノス	-0.263	0.141	-1.864	0.062
	サンフレッチェ広島	-0.131	0.137	-0.957	0.338
	FC東京	-0.051	0.141	-0.361	0.718
	名古屋グランパス	-0.287	0.146	-1.968	0.049
	ヴィッセル神戸	-0.273	0.144	-1.892	0.058
	アルビレックス新潟	-0.413	0.158	-2.609	0.009
	ヴァンフォーレ甲府	-0.524	0.169	-3.100	0.002
	ベガルタ仙台	-0.315	0.144	-2.180	0.029
対戦相手 チーム	浦和レッズ	0.012	0.158	0.074	0.941
	鹿島アントラーズ	0.015	0.156	0.093	0.926
	柏レイソル	0.245	0.153	1.599	0.110
	サガン鳥栖	0.178	0.151	1.179	0.238
	川崎フロンターレ	-0.012	0.149	-0.080	0.936
	横浜F・マリノス	-0.005	0.153	-0.031	0.975
	サンフレッチェ広島	0.004	0.152	0.023	0.982
	FC東京	0.026	0.154	0.169	0.866
	名古屋グランパス	0.312	0.142	2.202	0.028
	ヴィッセル神戸	0.236	0.144	1.644	0.100
	アルビレックス新潟	0.379	0.144	2.628	0.009
	ヴァンフォーレ甲府	0.274	0.149	1.839	0.066
	ベガルタ仙台	0.256	0.142	1.804	0.071
内的環境	午後1時	-0.346	0.382	-0.906	0.365
	午後2時	-0.111	0.370	-0.298	0.765
	午後3時	-0.283	0.374	-0.756	0.450
	午後4時	-0.317	0.376	-0.843	0.399
	午後5時	-0.650	0.428	-1.518	0.129
	午後6時	-0.201	0.374	-0.536	0.592
	午後7時	-0.317	0.368	-0.861	0.389
	午後8時	-14.926	542.357	-0.028	0.978
	観客数	-0.050	0.110	-0.452	0.651
	H/A×観客数	0.095	0.133	0.714	0.475
H/A	-1.008	1.298	-0.777	0.437	
外的環境	曇ダミー	-0.048	0.061	-0.785	0.433
	雨ダミー	-0.045	0.107	-0.421	0.674
	気温	0.155	0.091	1.716	0.086
	湿度	0.105	0.098	1.069	0.285
試合 データ	分析対象チーム・警告数	0.010	0.026	0.395	0.693
	対戦相手チーム・警告数	0.003	0.026	0.118	0.906
	分析対象チーム・退場1人目	-0.405	0.163	-2.480	0.013
	分析対象チーム・退場2人目	0.032	0.595	0.053	0.958
	対戦相手チーム・退場1人目	0.156	0.117	1.329	0.184
	対戦相手チーム・退場2人目	0.403	0.401	1.005	0.315
	分析対象チーム・シュート	0.073	0.008	9.449	0.000
	対戦相手チーム・シュート	0.011	0.008	1.450	0.147
	分析対象チーム・CK	-0.065	0.012	-5.330	0.000
	対戦相手チーム・CK	0.027	0.012	2.259	0.024
	分析対象チーム・FK	-0.006	0.007	-0.891	0.373
	対戦相手チーム・FK	0.016	0.007	2.168	0.030
	サンプルサイズ				
AIC					3079.9
疑似決定係数 R^2					0.192

得点が決まる要因として強く影響しているのが、分析対象のチームがシュートをした本数であり、シュートをすればするほど得点が決まりやすいということが分析から示された。一方で、分析対象のチームがコーナーキックにする本数が少なければ少ないほど得点が決まりやすいという結果が示された。これは、コーナーキックがどんなプレーなのかを確認する必要がある。コーナーキックとは、対戦相手のチームがボールをゴールラインから外に出した後に発生するプレーである。したがって、コーナーキックにする本数が少なければよいというのは、シュートをただ打つのではなく、しっかりとゴールに入れればコーナーキックにならないことを含んでいることが示されている。その他に影響している要因として、分析対象チームが 1 人の退場者を出してしまうことで、得点が決まりづらくなることが示された。

これらのことを踏まえ推定結果を用いることで、勝利するためのシミュレーションを行うことが出来ると考えられる。

4.おわりに

本研究では、ポアソン回帰分析により得点を決めるための要因が何であるかを検討することを目的とした。その要因としては、外的環境にあたる天候、気温、湿度、内的環境にあたる試合の開催時間、観客数、ホームかアウェイかの区別、試合データにあたる警告数、退場数、総シュート数、総コーナーキック数、総フリーキック数を取り扱った。その結果、シュートを放てば放つほど得点する可能性が高まるということが示された。また、分析対象チームがコーナーキックにすればするほど得点する可能性が低くなるという結果が現れたことに興味深く、しっかりとシュートを GK や DF に防がれることなく、枠内にシュートを放つことに意義があるということが示唆される。さらに、1 人でも退場をしてしまうと得点する可能性が低くなるという結果は、サッカーにおいて選手 1 人に対する責任が重く、その選手を交代出来ないことによるデメリットが如実に現れているものと考えられる。

そして、今回分析に取り扱ったデータは、2014 年から 2016 年までの 3 年間の J リーグが公式に公開されているデータのみで分析を行った。3 年間だけでなく 5 年、10 年分のデータを用いたり、全ての試合を映像で確認し一つ一つのプレーに対して評価を行ったり、選手一人一人から得られるデータであったりをも更なるデータの枠組みとして拡張してあげることで、より得点に起因する要因が判別することが出来ると考えられる。

さらに、得点の要因による期待値 λ_{ij} から、それぞれのクラブチームが対戦相手のクラブチームに、勝利する確率と負ける確率、引き分ける確率が求めることができると考えられる。

最後に、本研究で分析に取り扱ったのは J1 だけであったが、J2 や J3、JFL といった日本のほかのリーグだけでなく、その他のヨーロッパや南米のリーグなどに応用することが出来ることが考えられる。

謝辞

本研究は、サッカーを統計的な視点で分析をしてみたいと思ったのがきっかけとなって始めたものである。この卒業論文を書いていく上で、栗原由紀子先生と大橋忠宏先生のお二人から様々な助言を頂いた。そして、二年間ゼミナールが一緒であった叶さんと鷺尾さん、その他産業情報コース4年生とは励ましあいながら進めていくことが出来た。ここに記して感謝の意を表すものである。

参考文献

- 公益財団法人日本サッカー協会、審判、「ルールを知ろう」 <http://www.jfa.jp/referee/rule/>
Jリーグ公式サイト、「成績・データ」 <http://www.jleague.jp/stats/>
- 秋山裕 (2013), 「Rによる計量経済学」, オーム社
- 草野修治 (2012), 「サッカーにおけるシュート地点とゴールイン地点の傾向に関する分析 — 「2010年度 Jリーグディビジョン」1得点ランキングトップ10のJリーガーにおいて—」, 『仙台大学紀要』, 第44巻, pp.31-41
- 坂田幸繁・栗原由紀子 (2016), 「統計学」, 丸善プラネット
- 田村達也・堀野博幸・土屋純 (2015), 「サッカーにおけるボール奪取後の攻撃の分類法の提案と検討—2012年UEFAヨーロッパ選手権における速攻とポゼッション攻撃に注目して—」, 『スポーツ科学研究』, 第12巻, pp.42-55
- 西尾裕 (2011), 「J1リーグ優勝に関する統計的分析」, 『2010年度卒業論文要旨集』(南山大学)
- 福原崇之 (2007), 「サッカーJ1リーグにおけるホームアドバンテージの要因分析」, 『青山社会科学紀要』(青山学院大学), 第36巻第1号.
- 三家礼子・塚田将太・板橋智也・富田平・高橋卓 (2015), 「サッカーのビッグデータを用いた各ポジションの選手特性の検討」, 『人間工学』(早稲田大学), 第51巻特別号.
- 鷺崎早雄 (2011), 「ポアソン回帰を用いたJリーグの得点モデルの推定—サッカーの勝敗はどの程度確率的なのか—」, 『環境と経営』(静岡産業大学), 第17巻第1号.