

Jリーグとプレミアリーグにおける得点要因分析と予測モデルの提案

An Analysis of Scoring Factors and a Proposal for a Predictive Model in the J.League and the Premier League

2241055 佐伯 慶太

Keita SAEKI

指導教員 秋葉 知昭

In recent years, the number of people participating in soccer has increased significantly. Furthermore, the 2026 FIFA World Cup has attracted significant attention. In this study, propose an analysis and a predictive model to examine the factors influencing goal scoring. Based on these results, compared the accuracy of conventional methods and the predictive model.

1. 緒言

近年、サッカーの世界人口が大幅に増加して、2006年時点で2億6500万人[1]、現在まで総人口の増加に伴いさらにサッカーの競技人口が増えることが予測される。昨今では、第104回全国高等学校サッカー選手権大会の開催やサッカーFIFAワールドカップ2026の開催に伴い注目を集めている。サッカー競技においては戦術の高度化やデータ分析技術の発展により、試合内容を客観的に評価する試みが盛んに行われている。特に、勝敗に直結する得点は試合結果を左右する最も重要な要素の1つであり、その発生要因を明らかにすることは、戦術を組み立てることや選手育成において大きな意義を持つ。また、リーグによる違いや開催地等の違いにより得点要因が変わってくることも考えられる。これまでは、シュート数やポジション率、パス成功率などの試合スタッツデータと得点との関係について分析が行われてきた。また、近年では得点に至るプレーの連続性や空間的要因にも注目が集まっている。しかし、これらの研究の多くは特定のリーグや大会に限定されており、得点要因の一般化や複数の要因を統合的に検証した研究は十分とは言えない。鈴木らは、得点機会の分析において、パスプレーに着目した研究を行った[2]。そこで、本研究では、サッカーの試合データを用いて、得点に影響を与えと考えられる複数の要因を定量的に分析し、それらと得点の関係性を明らかにすることを目的とする。具体的には、シュートに関する指標や自チームがした反則の回数といった客観的なデータを用いて試合状況などに着目し、得点との関連性を分析する。

2. 分析対象と使用する変数

本研究では、データの対象として、プレミアリーグにおいては2023年6月から2024年5月まで

の1年間の試合を用いる。Jリーグにおいては2024年2月から2024年12月までの1年間の試合を用いる。合計1520試合分である。また、それぞれのリーグごとにホームゲームとアウェイゲームに層別して分析を行う。

目的変数：得点

説明変数：支配率、シュート数、カード枚数、コーナーキック数、ファール数、オフサイド数、XG、オッズ

3. 分析結果

変数間の相関を調べるため、相関分析を予備解析として行った。結果を抜粋して表1に示す。

表1 Jリーグ ホームゲームの相関分析の結果

No	変数名	支配率	シュート数	カード	コーナー	ファール数	オフサイド	XG	オッズ
2	支配率	1.000	0.189	-0.127	0.285	-0.221	-0.081	0.156	-0.214
3	シュート数	0.189	1.000	-0.140	0.369	-0.173	0.079	0.329	-0.195
4	カード	-0.127	-0.140	1.000	-0.143	0.236	0.047	-0.178	0.131
5	コーナー	0.285	0.369	-0.143	1.000	-0.138	-0.032	0.415	-0.214
6	ファール数	-0.221	-0.173	0.236	-0.138	1.000	-0.016	-0.191	0.179
7	オフサイド	-0.081	0.079	0.047	-0.032	-0.016	1.000	0.017	-0.076
8	XG	0.156	0.329	-0.178	0.415	-0.191	0.017	1.000	-0.315
9	オッズ	-0.214	-0.195	0.131	-0.214	0.179	-0.076	-0.315	1.000

StatWorks ver.5.82 を使い、Jリーグのホームゲーム、アウェイゲーム、プレミアリーグのホームゲーム、アウェイゲームで多変量解析をした。その結果の例を、表2に示す。

表2 Jリーグのホームゲームの重回帰分析の結果

目的変数名	重回帰係数	寄与率R ²	R ²	R ²	
得点		0.792	0.627	0.620	0.613
	残差自由度				
		374	1.179		
vNo	説明変数名	分散比	P値 (上側)	偏回帰係数	標準偏回帰
0	定数項	40.6375	0.000+		
2	支配率	0.8573	0.355	-0.004	-0.036
3	シュート数	60.1021	0.000	0.099	0.399
4	カード	0.7191	0.397	-0.045	-0.044
5	コーナー	2.0136	0.155	-0.035	-0.079
6	ファール数	6.7537	0.010	0.040	0.122
7	オフサイド	2.3626	0.125	0.066	0.075
9	オッズ	0.6191	0.432	-0.061	-0.036

P値の異常値がないため、結果として、採用した。以下より、解析結果で目立つ変数を抜粋して考察する。

まず、各リーグ、開催地で変数の種類が異なるため、最終的な変数の種類は少し異なる。標

準偏回帰係数の値が大きかったのは、シュート数であった。これは、点数を取るには積極的にゴールを狙うというサッカーにおいて基本的な原理が統計的に示唆された。また、コーナーキックの回数が得点に負の影響を与えることが見て取れる。コーナーキックによって、決めきることができなければコーナーキック取得側のチームは相手陣地に多くの人数を割いてしまっているため、カウンターによって逆に点数を決められる確率が上がることを示唆している。

4. 予測結果と比較からの考察

多変量解析から得られた偏回帰係数により、以下の予測式を作成した。

S_i

$$= a(1)m_i + b(1)n_i + c(1)o_i + d(1)p_i + e(1)q_i + f(1)r_i + g(1)s_i$$

i : 予測するリーグ、開催地

S_i : i の得点数

m_i : i の支配率

n_i : i のシュート数

o_i : i のカード枚数

p_i : i のコーナーキック数

q_i : i のファール数

r_i : i のオフサイド数

s_i : i のオッズ

$a(1), b(1), c(1), d(1), e(1), f(1), g(1)$ はそれぞれに対する偏回帰係数の値である。

算出した予測式の結果の一部を抜粋して表 3 に示す。

表 3 Jリーグのホームゲームの予測式の精度

予測値	実績値	残差	相対誤差	絶対値	MAD
0.92157	2	-1.07843	-1.17021	1.17021	0.65858452
1.39757	2	-0.60243	-0.43106	0.43106	
1.57120	1	0.57120	0.36354	0.36354	
1.67115	3	-1.32885	-0.79517	0.79517	
1.95702	3	-1.04298	-0.53294	0.53294	

MAD 値をみると、0.5 前後であるためある程度高い予測精度となった。

得られた結果の精度を MAD 値と RMSE 値で他手法と表 4 にて比較する。

表 4 他手法との比較

		J ホーム	J アウェイ	プレ ホーム	プレ アウェイ
本研究	MAD	0.658585	0.633203	0.473127	0.423398
移動 3 年	MAD	0.681431	1.210011	0.208334	0.240025
移動 5 年	MAD	0.531313	0.340079	0.398368	0.222222
本研究	RMSE	0.970139	1.287587	1.421068	0.783990
移動 3 年	RMSE	1.247221	1.556351	0.954521	0.447217
移動 5 年	RMSE	1.050714	0.558570	1.050714	0.521536

本研究の MAD による予測精度の分析では、どのリーグまたは開催場所においても、MAD が 0.5 点前後だったため比較的高い予測精度となった。また、開催場所を問わず Jリーグよりもプレミアリーグの方が高い予測精度を示した。特に、プレミアリーグのアウェイゲームが最小の値で、最も安定した予測ができていた。しかし、RMSE による分析では、1 点前後であったため MAD による評価に比べて予測精度が低くなった。移動平均法と本研究の値を MAD を用いて比較すると、開催場所を問わず Jリーグの予測精度が高い順に移動平均法 (5 試合)、本研究、移動平均法 (3 試合) となった。移動平均区間の長さの影響を受けた結果、5 試合で平均をとった場合よりは精度が落ちたものの、3 試合の場合よりは精度が高くなった。移動平均法と本研究の値を RMSE を用いて比較すると、リーグどうしの予測精度の順位付けは一概には不可能であった。しかし、Jリーグのホームゲームのみに着目すると、本研究が最も高い予測精度という結果となった。本研究の結果としてリーグや開催場所を問わず RMSE よりも MAD の方が高い予測精度となった。

5. 結 言

本研究では、サッカーの得点要因分析を行った。その手法として、相関係数行列を用いて説明変数の抽出を行ったのち、重回帰分析を行った。結果としては、リーグや開催場所を問わず「シュート数」が得点に特に強く影響していることが導かれた。また、「カードの枚数」や「コーナーキックの回数」が増えると得点機会が得られなくなるという結果になった。今後の課題としては、説明変数の増加や試合データの追加を行うことによって予測精度を上げることである。また、今回はリーグ分析を行ったが、ワールドカップといった国を代表するチームごとに分析を行うとまた違った結果が得られるであろう。

参 考 文 献

- [1] FIFA: https://rtmag.co.il/images/FIFA_BIG_COUNT.pdf
- [2] 鈴木: プロサッカーリーグにおける得点機会獲得のための攻撃プレーの分析: パスプレーの着目して: 体育学研究, 64, 761-764
https://www.jstage.jst.go.jp/article/jjpehss/64/2/64_19002/_article/-char/ja/