

J1 リーグの観客数に影響する要因の探索的研究と 予測モデルの構築

情 長岡 佳汰
指導教員 宋 財沄

1. はじめに

本研究の目的は、日本プロサッカーリーグ（J1 リーグ）における観客数に影響を与える要因を探索的に明らかにすること、そして精度の高い観客数予測モデルを構築することにある。プロスポーツ経営において観客動員数の把握は収益管理やマーケティング戦略の立案に直結する極めて重要な課題である。

本研究の背景には、学術的背景と社会的背景がある。学術的背景では、J リーグに関する観客数研究はメジャーリーグサッカー（MLS）やプレミアリーグと比較して少なく、海外研究で有効性が認められている変数が J リーグの文脈で十分に検証されていない。また、従来の国内研究では、多重共線性や過学習、誤差項の正規性といった統計的仮定の検証が不十分な例が見受けられる。社会的背景では、チケットの完売に伴う転売問題の発生により、需要予測に基づいたダイナミックプライシングの重要性が高まっている。また、J. LEAGUE は J リーグ露出戦略に取り組んでおり広告などのプロモーション活動の効果検証やどのような要因が観客数に影響するのか明らかにすることがあげられる。

2. 先行研究の概観

国内の先行研究では、チケット価格、チームの順位、天候、ダービーマッチ、開幕戦、平日・休日といった変数が観客数に影響を与えることが示されてきた（岩村他 2022）。海外の研究では、試合結果の不確実性や、スター選手の存在、スタジアムの築年数などが分析対象となっている。また、予測手法としては、重回帰分析のほかに、ニューラルネットワークやランダムフォレスト、近年では勾配ブースティング木（LightGBM 等）を用いた事例も報告されている。

3. データと分析手法

本研究では、2017 年から 2023 年までの J1 リーグ 1,604 試合をサンプルとした。ただし、新型コロナウイルス感染症の影響により入場制限が行われた 2020 年から 2022 年の期間は除外している。分析にあたっては、応答変数を観客数とし、チーム特性、試合環境、マクロ経済指標、カレンダー要因など計 151 個の説明変数を用意した。

要因分析には最小二乗法（OLS）を用いた線形回帰分析を採用した。モデルの推定にあたっては、対数変換した観客数を応答変数とする以下のモデルを構築した。

$$\log(\hat{y}_i) = \beta_0 + \beta_1 Distance_i + \cdots + \sum \beta_{ij} TeamA_{ij} + \sum \beta_{ij} Month_{ij} + \sum \beta_{ij} Year_{ij}$$

モデルを推定する際には、多重共線性のチェックを行い、VIF 値が 10 を超える変数を処理した結果、最終的に 108 個の変数を使用した。また、誤差項の正規分布、線形性、分散均一性、独立性、多重共線性の 5 つの前提条件やデータを学習用データと検証用データに分割し過学習を検証することでモデルの妥当性を確保した。

予測モデルの構築では、勾配ブースティング木である XGBoost、LightGBM、CatBoost や過学

習抑制のために ElasticNet を用いた線形モデルで予測モデルを構築した。データ分割は時系列 hold-out 法を用い、2017-2023 年（除コロナ禍）を訓練データ、2024 年データの前半を検証データ、後半をテストデータとした。精度評価指標には MAE、MAPE、RMSE を用いた。

4. 分析結果と考察

線形回帰分析の結果、正の影響を与える要因としては、ホームチームおよびアウェイチームの市場価値、地上波・BS 放送の有無、天候、ダービーマッチ、週末・祝日開催（Weekend）、J1 リーグの節ごと平均観客数、特定のホームクラブ（FC 東京、名古屋、新潟、横浜 FM、浦和、G 大阪など）などがあげられ、負の影響を与える要因としては、両チーム間の移動距離（Distance）18 時以降のキックオフ、ホームチームの順位（数値が大きいほど下位のため負の影響、チーム間の順位差があげられる。さらに、正の影響が大きい一部のクラブについてさらに分析する。浦和レッズは、ホームチームとしての係数が最も大きい。多重共線性により除外された変数との相関を確認したところ、シャトルバスの運営やプロモーション費用との強い相関が認められ、これらの取り組みが動員力に寄与している可能性が示唆された。FC 東京は、都道府県人口との正の相関が強く、移動距離（距離抵抗）が少ない人口密集地を本拠地とすることが優位に働いている。アルビレックス新潟は、J2 在籍期間が長いものの、J1 定着期間の長い札幌を上回る動員傾向にある。これは、県内に他のプロ野球球団が存在しないことや、充実したシャトルバス運営が要因として考えられる。

予測モデルの構築では、グリッドサーチでパラメータの最適化を行い各モデルのテストデータに対する予測精度を比較した結果、線形モデルが全ての指標で最も高い精度を示した。

表 1. チューニング後のモデル精度比較

モデル	MAE	MAPE	RMSE
XGBoost	4638.3994	0.2378	6750.9833
LightGBM	4684.2742	0.2362	6809.3700
CatBoost	4410.3984	0.2239	6511.8964
線形モデル	4303.6954	0.2262	6381.9459

一般にテーブルデータでは勾配ブースティング木（GBDT）が高い精度を発揮するとされるが、本研究のようなサンプルサイズ（1,604）が比較的小さく、かつ多様な外的要因に左右されるドメインにおいては、線形モデルの方が、汎化性能が高いことが示された。

5. 結論

本研究により、J1 リーグの観客動員にはチームの市場価値やカレンダー要因に加え、移動距離や放送有無といった多角的な要因が影響していることが再確認された。特に、浦和や新潟の事例から、プロモーションやアクセス面の充実が、地域人口などの構造的制約を克服する鍵となる可能性が示唆された。予測モデルにおいては、線形モデルが最も実用的であることが証明されたが、今後はさらなる精度向上のため、サンプルサイズの拡大や、今回考慮しきれなかった詳細なプロモーション・イベントデータの統合が課題として挙げられる。

6. 参考文献

岩村聡・吉田幸幸・仲澤眞・井上雄平（2022）「プロサッカーイベントの観客動員数の規定要因：2000 年から 2003 年までの期間における説明モデルの検討」『法政大学スポーツ健康学研究』13, 1-14