

Jリーグ順位予想

芝浦工業大学 数理科学研究会
辻村将吾

2016年11月4日

1 研究背景

私自身、サッカーに興味があり小さいときにサッカーをやっていたため、サッカーに関する研究がしたいと思い今回の研究テーマにした。確率統計などの知識を用いることによって、よりサッカーを観戦するのが楽しみになるのではないかと考えた。

1.1 データと順位の因果関係

2015年と2016年のデータを比較することによって、何か因果関係が表れてくるのではないかと考えた。Jリーグが一昨年から導入をしている「トラッキングシステム」というデータを元にどのような因果関係があるかなどを調べていった。また、様々な値の平均(\bar{x})とすると

$$\bar{x} = x_1 + x_2 + x_3 + \dots + x_n = \sum_{i=1}^n x_i$$

また、散らばり具合を表した分散

$$\delta^2 = \frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})^2$$

これの2乗を取った形の標準偏差

$$\delta = \sqrt{\frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})^2}$$

など基本的な計算方法を用いることによって様々な計算をしていき、どのような因果関係があるか調べていった。

1.2 因果関係

よく走るチームがよく降格圏(16位~18位)の中に入るといった不思議なデータを発見した。また、このデータによって試合を作るチームが優勝には有利。ファウル数の平均をとると上位陣になればなるほどファウル数が増えていくといった関係がみられた。フィジカル面で負けないというところで上位陣になるためには必要になってくると感じた。オフサイド数やプレーが切れた分数などにより「縦に速いチーム」や「時間を使うのがうまいチーム」などの各々のチームの特徴がみられた。

2 今後の課題

今回調査したリーグの年度は2年限定と短い期間だったため、データを取り考察するのが難しいと感じたためもっと年数を増やしていきたい。Jリーグのトラッキングシステムを導入したのが一昨年からだったために、データをもっと増やした10年や20年後に同じ調査をすればまた新しい発見がある可能性があるため、様々な観点からデータを増やした10年後などにも同じ調査をしていきたい。その際にはデータを分析するだけでなく度数分布を用いて確率密度関数 $f(x)(\geq 0)$ によってその関数が離散型 g

$$\sum_{i=1}^n f(x) = 1$$

または、その関数が連続型

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

なのか確率分布や単純な確率などを求めていき、より深くまで因果関係のあるなしを決めていけるとさらに精度のいい結果になるのではないかと考えた。また連続的確率分布の期待値 $E(X)$

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

を計算することによっても深く調査できるのではないかと感じた。今回はサッカーに限っての研究をしたが次回以降からは様々なスポーツの順位や結果を考察することを今後の課題にしていきたいと感じた。

参考文献

[1] <https://www.jleague.jp/stats/>