

J1 リーグ優勝に関する統計的分析

2007MI169 西尾裕

指導教員：木村美善

1 はじめに

私はサッカーが好きなので、よくテレビで見たりスタジアムに見に行く。そのなかで、どうしたら試合に勝つのか、それはチームによって違う。したがって、どのようにしたらJリーグで優勝ができる可能性が一番高いのか、また自分の好きなチームはどのようにすれば優勝への可能性が高くなるのか、ということをも2009年度のJ1全18チームの成績を用いて分析してみた。

2 データについて

2009年度のJ1リーグ全18チーム、ホーム&アウェイ各1試合ずつの計304試合分のデータを用いる。具体的には「週刊サッカーダイジェストJリーグ2009シーズン総集編」([3]参照)のなかから、クラスター分析には1年間の「ゴール」、「シュート」、「パス」、「キープス」、「クロス」、「ドリブル」、「タックル」、「クリア」、「ブロック」、「インターセプト」、「ファール」、「失点」、「セーブ」、「接触数」の全14項目。ロジスティック回帰分析には、1チーム全34試合の各試合ごとの「結果」、「ゴール」、「シュート」、「CK」、「オフサイド」、「スタジアム」、「天候1」、「天候2」、「気温」、「湿度」、「順位1」、「順位2」の全12項目を用いる。

3 クラスタ分析

まず、それぞれがどのようなチームなのか、特徴を調べるためにクラスター分析を用いた。最も明確なクラスターができるのでウォード法を使い、2009年度のJ1リーグの全18チームを分析した。次にJ1チームの平均の「J1平均」を加えた全19チームでの分析を行った。2種類のクラスター分析は、ほぼ同じ結果になったので、19チームのほうの結果を述べる。

3.1 2009年度分析結果

クラスター分析を行った結果、図1のようになった。19チームを5群に分けて、この図の左から順に第1群~第5群とした。第1群はディフェンスに力をいれているチーム、第2群は攻撃力が少し弱いチーム、第3群はパスが多いチーム、第4群はシュート決定率が高いチーム、第5群は全てにおいて平均的なチームとなった。

3.2 考察

クラスター分析を行った結果、上位3チームがそれぞれ違う群に入った。したがって、どの群でも優勝を狙える可能性があるかと判断できる。しかし、2009年度では、第2群に属しているチームは攻撃力が低いと判断され、実際の結果でも下位チームが多くなっていることから、優勝を狙うにはやや難しそうである。以上より、優勝を狙うには、安定した攻撃力が不可欠と思われる。

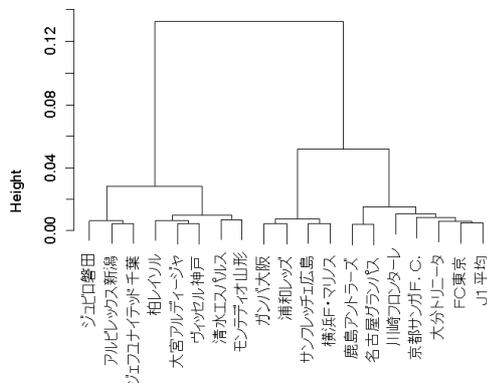


図1 2009年度デンドログラム

4 ロジスティック回帰分析

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取りうる値が[0,1]の外にでてしまい(-,)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([1]参照)。

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

本研究では目的変数 y を勝敗とし、11個の従属変数を用いて y を予測した。また「勝敗」、「スタジアム」、「天候1」、「天候2」、「順位1」、「順位2」にダミー変数を用いている。上位と下位の3チームずつの計6チームの分析を行ったが、スペース関係上、優勝チームの鹿島アントラーズについて説明する。

4.1 分析例:鹿島アントラーズについて

2009年度に1位だったチーム、鹿島アントラーズを例にとって分析を行ってみた。まずは全項目で回帰分析を行ってみた。次に最適なモデルを選ぶために増加法、減少法を用いて変数選択を行った。変数選択の結果は同じであった。最適なモデルは、「得点」、「スタジアム」、「天候2」、「湿度」の4変数を説明変数として用いたモデルになる。これをロジスティック回帰分析を用いて分析してみると以下のようなになる。なお「勝敗」は、勝ちなら1、それ以外を0、「スタジアム」は、ホームなら1、アウェイなら0、「天候2」は、雨なら1、それ以外なら0とするダミー変数である([2]参照)。

表 1 鹿島アントラーズの分析結果 (最適モデル)

	回帰係数	標準偏差	t 値	p 値
定数項	7.265	4.669	1.556	0.1197
得点	3.111	1.387	2.243	0.0249
スタジアム	3.123	1.839	1.698	0.0895
天候 2	3.760	2.340	1.607	0.1081
湿度	-18.470	8.918	-2.071	0.0383

分析結果を見てみると、湿度は負の方向に強く効いており、他の 3 項目は正の方向をに効いている。したがって、湿度が高いほど、勝率がさがっていくと考えられる。他の 3 項目はそれほど大差がないので、特にどれが強く効いているとはいえないだろう。また、全ての試合の勝率を先ほどの式を用いて計算した結果を表にまとめてみた。今回は勝率の高い順、低い順にそれぞれ 3 試合ずつのせることにした。なお、表の下にいくほど勝率は上がっていくようにしてある。上から順番に第 29 節、第 20 節、第 22 節、第 30 節、第 31 節、第 33 節である。

表 2 鹿島アントラーズの計算結果 1

勝率	得点	スタジアム	天候 2	湿度
0.00026	0	0	0	0.84
0.00031	0	0	0	0.83
0.00239	0	0	0	0.72
0.99996	3	1	1	0.72
0.99999	2	1	0	0.25
0.99999	5	1	0	0.53

計算結果を見てみると、得点が正に一番強く効いていると考えられる。負に効いているのが湿度のみなので、それ以外の条件が同じなら、湿度の高い方が勝率が低くなっている。得点が一番強く効くことはあたりまえなので、得点を抜いた 10 項目で増加法、減少法を行った結果、増加法では「オフサイド」、「スタジアム」、「天候 1」、「天候 2」、「湿度」の 5 項目となり、減少法では「オフサイド」を抜いた 4 項目となった。決定係数をみると、増加法の方がいいので、増加法を用いた変数選択の 5 項目を最適モデル 2 とした。なお「天候 1」は晴れなら 1 それ以外なら 0 とするダミー変数である。

表 3 鹿島アントラーズの分析結果 (最適モデル 2)

	回帰係数	標準偏差	t 値	p 値
定数項	8.7349	3.7714	2.316	0.0206
オフサイド	-0.5862	0.3652	-1.605	0.1085
スタジアム	2.9734	1.3454	2.210	0.0271
天候 1	2.1071	1.2711	1.658	0.0974
天候 2	5.3203	2.4215	2.197	0.0280
湿度	-14.1897	5.5295	-2.566	0.0103

得点を抜いた分析結果をみてみると、スタジアム、天候 1、天候 2 が正に効いて、オフサイド、湿度が負に効いている。回帰係数をみると、正に一番効いているのは天候 2 で、負に一番効いているのは湿度と予想できる。この結果を用いて先ほどの式を使い、最適モデル 1 と同じように結果をのせてみた。なお上から順番に第 29 節、第 20 節、第 26 節、第 30 節、第 11 節、第 31 節である。

表 4 鹿島アントラーズの計算結果 2

勝率	オフサイド	スタジアム	天候 1	天候 2	湿度
0.0006	7	0	0	0	0.84
0.0258	1	0	0	0	0.83
0.0477	5	0	0	0	0.62
0.9886	4	1	0	1	0.72
0.9960	1	1	1	0	0.62
0.9997	3	1	1	0	0.25

計算結果をみてみると、正に一番効いているのは天候 1 で、負に一番効いているのは湿度であると判断できた。回帰係数的には、天候 1 のが天候 2 より小さいが、晴れよりも雨の方が湿度が高くなるので、天候 2 より天候 1 の方が値が大きくなる。したがって天候 1 の方が強く効くと判断した。

4.2 考察

湿度が負に強く効いているので、湿度が高いと勝率が低くなる。湿度が高くなるとなにか悪いのか考えると、いつも以上に体力が奪われてしまう。したがって、体力勝負の試合に弱いと考えられる。このチームは体力がある選手が少ない、もしくは体力が奪われてチーム力が落ちてしまうと判断できた。他の変数もみてる。得点がある場合とない場合を比べると、ない場合のほうが天候 2 が正に強く効くことがわかった。したがって、このチームの場合、得点が一番正に効き、次に天候が強く効いている。よって、天気あまり左右されないチームなのだとわかった。ちなみに試合前の条件として最も勝率が高くなる場合は、ホームの試合で、天気は晴れ、できるだけ湿度が低い場合である。

5 おわりに

本研究を終えて、やはり得点が一番勝率に関係していて、次に関係しているのは湿度であった。これは自分の予想外のものであり、これから試合を見ていくなかで、また違う視点から見るので楽しみである。本研究では、18 チーム中 7 チームしか分析できなかったが、残りの 11 チームも同様な結果となるのか確かめるのが今後の課題である。

参考文献

- [1] 中村永友:多次元データ解析法, 共立出版, 2009
- [2] 丹後俊朗・高木晴良・山岡和枝:ロジスティック回帰分析 - S A S を利用した統計解析の実際 -, 朝倉書店, 1996
- [3] 週刊サッカーダイジェスト J リーグ 2009 シーズン総集編, 2010