

サッカーW杯における 勝敗に関する統計的分析

2011SE278 都築直人

指導教員：白石高章

1 はじめに

本年、第20回サッカーW杯ブラジル大会が開催された。4年に一度のW杯を多くの人々が楽しみにしている。観戦するのにあたり、どのようなことが勝敗に関係するのかを知りたいと思った。それを知ることで普段漠然と見ていた試合が、より様々な角度から試合を楽しめると考えた。また、サッカーを見ない人でもどのようなことが勝敗に関係するのかを知れば、サッカーに興味を持ち、試合を楽しめるのではないかと考えた。

では、サッカーの試合でどのようなことが勝敗に関係するのだろうか。日本が出場した1998年フランス大会、2002年日本韓国大会、2006年ドイツ大会、2010年南アフリカ大会、2014年ブラジル大会のデータをもとに解析していく。

2 データと考察

集めた統計データをもとに、考察について述べる。

2.1 初戦の結果によるグループリーグ通過について

W杯5大会での、初戦の結果によってグループリーグを通過する確率について見ていく。

表1、表2はW杯5大会のデータを参考文献[1]を介して集めた。

表1 W杯5大会の初戦の結果によるグループリーグ通過・敗退について

試合数	勝ち	引き分け	負け	合計
初戦の結果	60	40	60	160
割合	0.375	0.250	0.375	1.000
GL突破	51	22	7	80
割合	0.638	0.275	0.088	1.000
GL敗退	9	18	53	80
割合	0.113	0.225	0.663	1.000

表2 初戦の結果でグループリーグを通過する割合を示したデータ

試合数	勝ち	引分	負け	合計
初戦	60	40	60	160
GL突破	51	22	7	80
割合	0.850	0.550	0.117	0.500

この表1、表2より、初戦を勝った場合は、通過率が8割でとても高い値である。初戦を負けた場合は、1割しか通過できない。

2.2 ブラジル大会優勝国ドイツのW杯での得失点について

表3 W杯のドイツの得失点の頻度をまとめたデータ

得点	頻度	失点	頻度	
0	1	0	10	
1	4	1	2	
2	5	2	3	
3	1	3	0	
4	3	4	0	
5	0	5	0	
6	0	6	0	
7	0	7	0	
8	1	8	0	
計	37	15	8	15

平均得点は2.467、

平均失点は0.533である。

2.3 日本のW杯での得失点について

表4 W杯の日本の得失点の頻度をまとめたデータ

得点	頻度	失点	頻度	
0	5	0	5	
1	7	1	4	
2	2	2	3	
3	1	3	1	
4	0	4	2	
5	0	5	0	
6	0	6	0	
7	0	7	0	
8	0	8	0	
計	14	15	21	15

平均得点は0.933、

平均失点は1.400である。

3 データ解析の結果

3.1 初戦の結果によるグループリーグ通過率

参考文献[2]、[3]を用いてW杯5大会における、初戦の結果でグループリーグ通過率に関しての解析を表2をもとに行った。

A:初戦に勝ち、1-1/30の確率で予選を通過する事象

$$= \{p_1 \in I_1\}$$

B:初戦に引き分け, 1-1/30 の確率で予選を通過する事象

$$= \{p_2 \in I_2\}$$

C:初戦に負け, 1-1/30 の確率で予選を通過する事象

$$= \{p_3 \in I_3\}$$

・ A の場合

$$\frac{L}{K \cdot F_L^K(\frac{\alpha}{6}) + L} < p < \frac{K^* \cdot F_{L^*}^{K^*}(\frac{\alpha}{6})}{K^* \cdot F_{L^*}^{K^*}(\frac{\alpha}{6}) + L^*}$$
 の式に,

$$K = 20, L = 102, K^* = 104, L^* = 18, \alpha/6 = 0.0167$$

を代入し計算すると,

$$0.7242 < p_1 < 0.9934$$

B と C も同様に計算すると,

$$0.3724 < p_2 < 0.7188$$

$$0.0444 < p_3 < 0.2354$$

A,B,C が信頼係数 0.9 の同時信頼区間 $p_1 \in I_1, p_2 \in I_2, p_3 \in I_3$ が予選を通過する比率に対する信頼係数 0.9 の同時信頼区間である.

これにより, 初戦を勝ってグループリーグを通過する確率は 72% 以上, 引き分けてグループリーグを通過する確率は 37% 以上 71% 以下, 負けてグループリーグを通過する確率は 24% 以下である.

3.2 ドイツと日本の W 杯での得失点について

W 杯での得点分布と失点分布がいずれもポアソン分布 $P(\mu)$ に従うと仮定し, 母平均 μ の最尤推定値を求める.

参考文献 [2], [4] を用いて, W 杯での得失点についての解析を, 表 3, 表 4 をもとに行った.

・表 3 の得点の場合

全データの推定値は $37/15=2.467$ である. 1 得点以上のデータは $m=14$ 個あり, 平均値は $\bar{x} = 2.643$ である. これより, 最尤推定値を求める反復法は,

$$\mu^{(t+1)} = 2.643(1 - \exp[-\mu^{(t)}]) \text{ となる.}$$

$$\mu^{(0)} = 2.643 \text{ として反復させると,}$$

表 5

ITE	μ
0	2.642
1	2.454
2	2.415
3	2.406
4	2.404
5	2.403
6	2.403

表 5 のような推移をたどり, $\hat{\mu} = 2.403$ に収束する.

次に参考文献 [5] より, McKendrick の方法を適用する. $m=14$ および $\omega = 37$ であるので, 反復法は,

$$f_0^{(t+1)} = (14 + f_0^{(t)}) \exp[-\mu^{(t)}]$$

$$\mu^{(t+1)} = \frac{37}{(14 + f_0^{(t+1)})}$$

であり, $\mu^{(0)} = 2.643$ および $f_0^{(0)} = 0, n^{(0)} = 14$ から始めて, 推移の結果, $\hat{\mu} = 2.404$ が得られ収束する. この場合の全データ数 n の推定値は 15.39 と実際の値 15 に極めて近い. また参考文献 [6] より, Moore の非反復推定法では, 得点 2 以上のみの観測値の和は 33 であるので, 推定値 $\hat{\mu} = 33/14 = 2.357$ を得る.

・表 3 の失点の場合

同様に計算すると, 0.529 が得られ収束する.

・表 4 の得点の場合

同様に計算すると, 0.851 が得られ収束する.

・表 4 の失点の場合

同様に計算すると, 1.981 が得られ収束する.

表 3 の失点と表 4 の詳しい計算は本論に記載されている.

4 おわりに

本研究を通して, サッカー W 杯における勝敗に関する統計的分析を進めてきた. 2.1 節と 3.1 節より初戦を勝ってグループリーグを通過する確率と初戦を負けてグループリーグを通過する確率にはとても大きな差があることが分かった. 2.2 節, 2.3 節より, 日本とドイツでは得失点の頻度や平均得失点の両方に大きな差があることが分かった. 3.2 節より最尤推定値を求め, 日本とブラジル大会優勝国ドイツでは大きな差があることが分かった. これらの差を縮めることが日本が今後 W 杯で勝つことに繋がると考えられる.

参考文献

- [1] ワールドカップのデータベース【WorldCup's world】
<http://members.jcom.home.ne.jp/wcup/>.2014.12.
- [2] 白石高章: 『統計科学の基礎-データと確率の結びつきがよくわかる数理』. 日本評論社, 東京, 2012.
- [3] 藤井良宜: 『R で学ぶデータサイエンスカテゴリカルデータ解析』. 共立出版, 東京, 2010.
- [4] 岩崎学: 『カウントデータの統計解析』. 朝倉書店, 東京, 2010.
- [5] McKendrick, A.G (1926) Applications of mathematics to medical problems. Proceedings of the Edinburgh Mathematical Society, 44, 98-130.
- [6] Moore, P.G. (1954) A note on truncated Poisson distribution. Biometrics, 10, 402-406.