

世界の野球の強さに関する統計的分析

2011SE080 石原裕久

指導教員：松田真一

1 はじめに

私は、野球が好きでテレビでもみており、最近ではWBCという大きな大会をみた。その大会では日本が負けて悔しい思いをした。そこでそれぞれのチームは、どのような特徴で、強さの要因は何なのかなど様々な疑問について統計学を用いて分析することができないだろうかと考えたことがこのテーマを選んだきっかけである。

2 データについて

日本野球機構 [5] から「WBC の勝敗」のデータを用いた。また、強さの推定値では鬼頭・高田の卒論 [1] を用いて分析を行った。重回帰分析では、Web[3] から各国ごとの1シーズンの試合数、「国内リーグチーム数」、「国際大会開催数」のデータを用い、日本野球機構 [5] から「WBC の平均防御率」と「平均打率」を、Web[4] から「100万人当たりの殺人件数」、「1人当たりのGDP」を用いた。主成分分析では日本野球機構 [5] から各国ごとのチーム成績のデータを用いた。チーム成績は、「打率」、「得点」、「長打」、「盗塁」、「出塁率」、「三振」、「防御率」、「被安打」、「与四死球」、「奪三振」、「失点」を用いた。

3 分析方法

Bradley-Terry モデル (以下では BT モデル) (竹内・藤野 [4] 参照)、重回帰分析、主成分分析、クラスター分析、t 検定を用いた。

4 BT モデルの推定値

| | | | | |
|---------|-------|-----|---------|-------|
| 1, 日本 | 138.1 | 8, | キューバ | 76.8 |
| 2, 中国 | 20.1 | 9, | オランダ | 42 |
| 3, 台湾 | 47.7 | 10, | プエルトリコ | 77.1 |
| 4, 韓国 | 147.7 | 11, | オーストラリア | 4.08 |
| 5, カナダ | 26.3 | 12, | ドミニカ | 130.1 |
| 6, メキシコ | 14.6 | 13, | イタリア | 19.8 |
| 7, アメリカ | 39.5 | 14, | ベネズエラ | 64.1 |

1番高いのは韓国である。韓国は過去三大会で一番日本に勝っている国であるので韓国が一番高いという結果になったのではないかと思う。

5 カモ・苦手関係

過去3大会分の勝敗と各チームごとの対戦数、各対戦相手の勝利数のデータをもとにカモ・苦手関係を出した。目立ったカモ・苦手関係はないが日本、アメリカ、韓国の三国では日本はアメリカに強いといえるが、韓国相手には弱い。アメリカは韓国相手には強いといえるが日本相手には

弱い。日本はアメリカ相手には強いが韓国相手には弱いという三すくみの関係になっていることがわかる。総合的には国際大会ではカモ・苦手は発生しにくいのではないかと思う。

6 重回帰分析

BT モデルで出した強さを目的変数として重回帰分析を行った。また、残差分析を行ったところベネズエラはクック距離1を外れたが、ベネズエラをはずしても正規性は乱していないと判断できデータに誤りがないと分かった。VIF を調べたところ、「チーム数」は10以上を超えていたのでチーム数を除いて分析した。決定係数は0.966で自由度調整済み決定係数は0.931であった。

表1 強さの推定値に対する重回帰分析結果

| | 係数 | 標準誤差 | t 値 | P 値 |
|-------------|---------|--------|--------|--------|
| (Intercept) | 22.88 | 47.61 | 0.481 | 0.6477 |
| シーズン | 0.2217 | 0.0971 | 2.283 | 0.0625 |
| 殺人 | -0.784 | 0.2502 | -3.136 | 0.0989 |
| 打率 | 0.476 | 0.1839 | 2.592 | 0.0411 |
| 防御率 | -19.84 | 1.770 | -11.21 | 0.0002 |
| GDP | -0.0449 | 0.0103 | -4.354 | 0.0048 |
| 開催数 | 2.484 | 1.884 | -3.441 | 0.0137 |

国際大会を開催するぐらい野球に力をいれたりまた、シーズンの試合数が多いほど野球に力をいれているということなので「シーズン」と「開催数」は、+にはたらいいたのではないかと思う。また、「殺人」に関しては治安が悪い国は発展途上国が多いので野球は道具の量が多くお金がかかるスポーツなので-にはたらいいたのではないかと考えられる。

7 攻撃面について

各国の特徴を分析するために、WBC 各三大会分のチーム成績を合算して1試合あたりの数値にすべて直した。準決勝以上進出経験のある8チームの攻撃面と守備面の成績の特徴を主成分分析とクラスター分析を用いて分析した。なお、攻撃面と守備面のクラスター分析は紙面の都合上割愛させていただく。

7.1 主成分分析

累積寄与率 86 % の第二主成分までを考察する。

- 第一主成分 (寄与率 0.701) 三振は正の値をとり、他はすべてマイナスの係数をとっている。三振数は他の変

数と逆にむけば優秀といえるので、「総合力」とした。

- 第二主成分 (寄与率 0.168) 打率と長打と三振が正の値をとり、盗塁が負の値をとっているので「ランナーの出し方に関する軸」とした。

8 守備面について

8.1 守備面の主成分分析結果

累積寄与率 89 %の第二主成分までを考察する。

- 第一主成分 (寄与率 0.753) 奪三振を除きすべてマイナスの係数であるため「総合力」とした。
- 第二主成分 (寄与率 0.145) 与四死球が大きくマイナスで、被安打がプラスであるので「ランナーの出し方に関する軸」とした。

9 攻撃・守備のクラスター分析

※ () 内は BT モデルでだした強さの推定値。

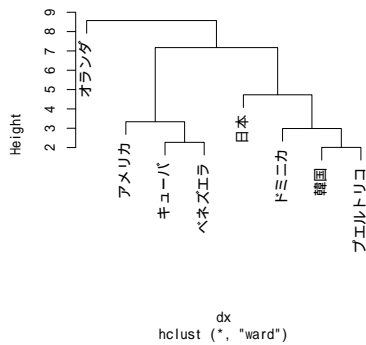


図 1 攻守総合のデンドログラム

- 第一群 オランダ (42)
投打でレベルアップが必要であると感ぜられる群。
- 第二群 アメリカ (39.5) キューバ (76.8) ベネズエラ (64.1)
打はよいが投手力が悪く、投打のバランスがとれていない群。
- 第三群 日本 (138.1)
投打のバランスがとれている群。
- 第四群 ドミニカ (130.1) 韓国 (147.1) プエルトリコ (77.1)
投手力は高いが打撃力が芳しくない群。

9.1 t 検定

攻撃・守備のクラスター分析で第二群と第三、四群で二標本 t 検定を行った結果、p 値=0.028 と棄却されるので、打撃力よりも投手力を向上させたほうが強さの上昇につながる事がわかる。

10 各国の特徴

ここでは日本、アメリカのみとさせていただきます。

10.1 日本

BT モデルの強さの推定値では 2 番目に高い数値を記録した。主成分分析では日本の攻撃面では、総合力の高いチームの中でも唯一盗塁を多く絡めて点をとっていく特徴がある。その反面打線の長打力に欠けており、大きいのが打てる打者が何人か必要であると感じた。守備面では、総合力が 2 番目に高く、四死球が少ない。一番バランスがとれている国ではないかと思う。

10.2 アメリカ

BT モデルの強さの推定値では高い値を示さなかった。攻撃 (長打) に特化したチームである。主成分分析では総合力の高いチームの中で長打力が圧倒的に高い。しかし、結果が伴っていないので盗塁をからめた攻撃をすると得点力アップが望めると思う。守備面では、全チームの中でも打たれてランナーを出すことが多いので、三振をとれる決め球の多い投手が必要ではないかと思う。

11 まとめ

本研究を通して、推定値に関しては 2 回優勝した日本が一番であると予想していたが実際は韓国が一番であった。重回帰分析では、野球以外の変数も強さに関係してくることもわかった。また、それぞれの国の特徴をみて、打撃面では総合力が高いチームでパワーと足のどちらかというところで対極化した。守備面では、ランナーを四死球で出すか、打たれて出すかというところがポイントであった。総合力が高いチームは被安打でランナーを出すことが多いので、四死球が多い投手はチームに入れるべきではないと思われる。最後の t 検定では打力よりも投手力を優先すべきかということがわかった分析であった。

12 おわりに

本研究では主に分析したのは 8 ケ国であった。野球自体競技人口が多くないので国数自体が少ない分析になってしまった。今後、WBC に参加する国が増えていくことを願いたい。

参考文献

- [1] 鬼頭薫・高田涼子:「S-plus における強さの推定」南山大学経営学部情報管理学科卒業論文, 1999.
- [2] 鈴木賢志: 鈴木賢志の研究室ページ, <http://www.dataranking.com/index.cgi?LG=j> (2014 年 10 月 20 日現在)
- [3] 世界ランキング統計局: <http://10rank.blog.fc2.com/blog-entry-214.html> (2014 年 10 月 20 日現在)
- [4] 竹内啓・藤野和建:「スポーツの数理科学」, 共立出版, 1988.
- [5] 日本野球機構: <http://www.npb.or.jp/wbc/> (2014 年 10 月 20 日現在)