

文章编号:1000-1735(2017)02-0145-05

DOI:10.11679/lxxblk2017020145

基于 Bradley-Terry 模型的中国男子篮球职业联赛球队实力分析

高长玉¹, 李 丹², 侯 文², 高清宇³

(1. 大连医科大学 现代教育技术中心, 辽宁 大连 116044; 2. 辽宁师范大学 数学学院, 辽宁 大连 116029;
3. 大连外国语大学 孔子学院工作处, 辽宁 大连 116044)

摘 要:随着体育事业及体育科学的发展,与体育相关的数据分析和统计工作也越来越得到重视. 基于 Bradley-Terry 模型,对 2015—2016 赛季中国男子篮球职业联赛数据进行分析,得到了 20 支参赛球队的实力指数估计值和主场效应指数,基本反映了各支球队真实的实力水平. 此外,对篮球技术指标进行因素分析,得到篮板球、助攻和投篮命中率是影响球队实力指数的主要技术指标,为今后预测体育赛事和提高球队的技战术水平提供依据.

关键词:成对比较数据; Bradley-Terry 模型; 实力指数

中图分类号: O213

文献标识码: A

近几年,随着我国政府对体育事业发展的支持力度不断加大,特别以中国男子篮球职业联赛(以下简称 CBA)为代表的职业体育赛事也得到了极大的发展,运动竞技水平也在不断提高,比赛精彩纷呈,吸引了广大体育爱好者的关注. 同时,与体育赛事相关的体育博彩业也得到了相应的发展. 当 2 支球队进行比赛,球队的实力往往决定比赛的结果,当一个强队同一个小队进行比赛时,我们理所当然会认为强队赢或者得分率会高,机会只会个别场次比赛中起或多或少的作用,这些都可以通过对比赛数据的统计分析来验证. 因此,对体育赛事结果进行数据分析是体育博彩业的基础.

Bradley-Terry 模型最早是由 Bradley 和 Terry^[1]在 1952 年提出的,是针对成对比较数据的模型. 后来,Bradley^[2-3]又详细地讨论了 Bradley-Terry 模型中的极大似然估计的渐近性、假设检验、置信空间等问题. 由于 Bradley-Terry 模型在成对比较数据模型中具有良好的统计性质,因此,Bradley-Terry 模型已应用到了生物医学、经济学、社会学等很多领域^[4-7]. 而很多体育项目比赛结果数据属于成对比较数据,Bradley-Terry 模型也可应用其中. 例如, Koehler 和 Ridpath^[8]利用 Bradley-Terry 模型分析了 1977—1978 赛季美国男子篮球职业联赛球队的实力,并评估出每支球队的主场优势. Mchale 和 Morton^[9]利用 Bradley-Terry 模型对 2000—2008 赛季的国际男子职业网球巡回赛选手的排名进行了预测. Balreira、Miceli 和 Tegtmeyer^[10]运用 Bradley-Terry 模型对 2002—2013 赛季的美国职业橄榄球比赛数据进行了拟合,给出了各支球队获胜的概率. 我国在这方面研究的文献很少.

本文采用 Bradley-Terry 模型拟合 2015—2016 赛季的 CBA 的赛事数据,估计各支参赛球队的实力指数,以及球队实力的影响因素分析.

收稿日期:2017-01-20

作者简介:高长玉(1963-),男,辽宁营口人,大连医科大学教授研究员级高级工程师.

1 Bradley-Terry 模型简介

1.1 Bradley-Terry 模型

设有 n 个个体参与成对比较,记为“1,2,⋯,n”,假设个体 r 有真实的能力参数 γ_r ,令 $P_{r,s}$ 表示个体 r 优于个体 s 的概率,Bradley-Terry 模型将 $P_{r,s}$ 表示为 Logit 模型形式

$$P_{r,s} = \exp(\gamma_r - \gamma_s) / [1 + \exp(\gamma_r - \gamma_s)]. \quad (1)$$

$P_{r,s}$ 只依赖 γ_r 与 γ_s 之差,考虑到模型的可识别性,模型的约束条件为 $\sum_{k=1}^n \gamma_k = 0$,或者某个 $\gamma_k = 0$, $k \in \{1, 2, \dots, n\}$.

在体育比赛中,不失一般性,比赛结果有 2 种情况. 设 Y_n 表示主队 a_r 与客队 a_s 的比赛结果, $Y_n = 1$, 表示主队 a_r 获胜, $Y_n = 0$, 表示客队 a_s 获胜. 如果考虑比赛的主场因素,就需要在 Bradley-Terry 模型 (1) 中加入所有球队主场效应参数 η , 模型表示为

$$P(Y_n = 1) = \exp(\eta + \gamma_r - \gamma_s) / [1 + \exp(\eta + \gamma_r - \gamma_s)]. \quad (2)$$

1.2 模型的解释变量

为了能够解释一些个体优于其他个体的原因,就需要找出可能影响个体能力差异的一些因素,然后建立回归模型,这些因素就是个体能力的解释变量.

以 γ_r 表示个体 a_r 的能力参数,如果每一个个体的能力都与解释变量 $x_{r1}, x_{r2}, \dots, x_{rm}$ 有关, $\beta_1, \beta_2, \dots, \beta_m$ 是解释变量的系数, U_r 是独立的误差项,并且 $U_r \sim N(0, \sigma^2)$. 个体能力 γ_r 与各解释变量的模型为

$$\gamma_r = \sum_{i=1}^m \beta_i x_{ri} + U_r.$$

个体 a_r 与个体 a_s 的能力值之间差异的模型为

$$\begin{aligned} \gamma_r - \gamma_s &= \sum_{i=1}^m \beta_i x_{ri} + U_r - \left(\sum_{i=1}^m \beta_i x_{si} + U_s \right) = \\ &= \sum_{i=1}^m \beta_i x_{ri} - \sum_{i=1}^m \beta_i x_{si} + U_r - U_s. \end{aligned} \quad (3)$$

将式 (3) 代入式 (1) 就构成一个广义线性混合模型,可以用惩罚拟似然方法对模型参数进行估计^[11].

2 CBA 球队实力及其影响因素分析

CBA 联赛开始于 1995 年,创办之初有 12 支球队参赛,发展到现在有 20 支球队参赛. 赛季比赛分常规赛和季后赛 2 个阶段进行,每年联赛最后 2 名降级,下级联赛的前 2 名升入 CBA 联赛. 在常规赛中,所有参赛球队进行双循环比赛,即每支球队要打 38 场比赛,分别在主场和客场各进行 19 场比赛,所有比赛的总场数为 380 场. 比赛计分方式为胜 1 场得 2 分,负 1 场得 1 分,弃权得 0 分. 常规赛按比赛胜场数/负场数比率确定常规赛总排名,胜场率高者名次列前. 2015—2016 赛季参赛的 20 支球队按它们在常规赛的积分顺序列于表 1 的第 1 列和第 2 列.

近几个赛季,各支球队为营造主场球迷氛围,球队在主场实力都得到充分展示,主场胜率较高,所以用带有主场效应的 Bradley-Terry 模型 (2) 拟合 2015—2016 赛季常规赛 380 场比赛结果数据. 由模型识别条件,设定八一队实力指数为 0,其他各队实力指数的估计值及其对应的标准误分别列于表 1 的第 3 列和第 4 列. 由于需要比较各支球队实力间的差异,故要计算拟标准误,其定义和计算方法见 Firth 和 Menezes^[12] 的论著,主要用于分类数据中类别之间比较. 各支球队拟标准误的值见第 5 列. 根据实力参数的估计值及其标准误计算得到的检验统计量 Z 值列于第 6 列,其对应的概率在第 7 列,根据各支球队实力估计值排名见第 8 列.

表 1 2015—2016 赛季 CBA 联赛球队实力估计结果

Table 1 Estimations of the ability about China Basketball Association in 2015—2016

球队	积分	实力估计值	标准误	拟标准误	Z 值	Pr(> Z)	实力排名
辽宁	69	3.8863	0.6471	0.3995	6.006	0	2
新疆	68	3.4761	0.6376	0.3935	5.452	0	3
四川	68	4.2481	0.6671	0.4277	6.368	0	1
广东	66	3.3957	0.6341	0.3878	5.355	0	4
山东	66	3.1064	0.6349	0.3954	4.893	0	5
广厦	65	3.0258	0.6342	0.3961	4.771	0	6
北京	61	2.4790	0.6107	0.3680	4.059	0	7
浙江	60	2.2880	0.6107	0.3733	3.747	0.002	8
深圳	57	1.9695	0.6083	0.3811	3.238	0.016	9
山西	56	1.7794	0.6067	0.3807	2.935	0.003	11
江苏	56	1.8632	0.6073	0.3817	3.073	0.002	10
上海	56	1.5794	0.6063	0.3788	2.605	0.009	12
福建	54	1.5598	0.6040	0.3858	2.582	0.010	13
青岛	54	1.4198	0.6040	0.3858	2.351	0.019	14
北控	51	1.0412	0.6052	0.4001	1.720	0.085	15
同曦	50	1.0405	0.6071	0.4075	1.714	0.087	16
吉林	48	0.5565	0.6144	0.4273	0.906	0.363	17
天津	46	0.0235	0.6282	0.4561	0.037	0.968	18
八一	45	0	—	0.4754	—	—	19
佛山	44	-0.2125	0.6530	0.4994	-0.325	0.667	20

注：数据来源于文献[13]

各支球队实力估计值显著性,可以从对应的检验概率看出. 主场效应参数是 0.0371(0.0068), 检验显著,说明 CBA 比赛具有主场优势效应. 如果要比较球队之间的实力差异,以辽宁队为例,它与新疆队实力估计值的差为 $3.8863-3.4761=0.4102$,其对应的拟标准误为 $\sqrt{0.3995^2+0.3935^2}=0.5607$,比值为 $0.4102/0.5607=0.7316$,可以认为两队实力没有显著性差异. 按此方法计算,辽宁队与北京队两队实力在 0.05 检验水平下有显著性差异,因此,可以认为辽宁队的实力明显高于积分排名低于北京队的各支球队,而与积分排名高于北京队的各支球队没有差异. 各球队的常规赛排名与实力估计值排名之间的 Kendell 秩相关系数为 0.968,在 0.01 水平下,两者排名相关性非常显著.

球队的实力往往决定比赛的结果,那么影响球队实力的因素是什么呢? 根据叶庆晖和邓飞^[14]以及张斌^[15]对篮球技术特点的研究文献,选择可能影响球队实力的 7 项技术指标代入模型(3)中进行影响因素的分析,具体技术指标及其参数估计结果见表 2.

广义线性混合模型中包括固定效应和随机效应 2 部分. 在随机效应检验中,剩余标准差为 0.8911(0.2604),非常显著,说明混合模型是成立的. 固定效应分析结果见表 2. 在表 2 中,固定效应中

显著影响球队实力的技术指标有投篮命中率、篮板球和助攻 3 项,其他技术指标对球队实力没有显著性影响.

表 2 影响因素估计结果

Table 2 Estimations of technical indicators factors affecting team's ability index

技术指标	回归系数	标准误	Z 值	Pr(> Z)
投篮命中率	6.6953	3.7934	1.765	0.039
篮板球	0.3271	0.1416	2.311	0.021
助攻	0.3501	0.1735	2.017	0.044
抢断	0.2292	0.2544	0.901	0.368
失误	-0.0197	0.1359	-0.145	0.885
盖帽	0.0424	0.0701	0.606	0.546
犯规	-0.0235	0.0292	-0.805	0.379

3 讨 论

基于 Bradley-Terry 模型分析了中国职业体育比赛中具有影响的 CBA 联赛赛事数据,探讨了参赛球队的实力指数及其影响因素,该方法也可以用于其他项目的赛事数据分析.

对 CBA 球队实力影响较大的常规技术指标分别是总投篮命中率、篮板球和助攻,这为球队有针对性训练,提高球队实力提供参考依据.不过本研究只局限于一个赛季数据,影响球队实力的技术指标只是在本赛季是显著的,在其他赛季这种影响是否显著,是否还有其他技术指标对球队实力也有影响还需进一步研究.同时,在不同国家联赛球队的技战术水平和特点不同,影响球队实力的技术指标也会有所差异.

一般运营成熟有影响的体育赛事赛季时间都比较长,各支球队的实力会受伤病、疲劳、客场的地域环境等各种因素的影响,球队实力会产生波动.如何能在基本 Bradley-Terry 模型中附加一些参数,能够反映这些因素造成的球队实力变化,是需要研究的课题.

参考文献:

[1] BRADLEY R A, TERRY M E. Rank analysis of incomplete block designs I. The method of paired comparisons[J]. Biometrika, 1952, 39(3): 324.

[2] BRADLEY R A. Incomplete block rank analysis; on the appropriateness of the model for a method of paired comparisons[J]. Biometrics, 1954, 10(3): 375-390.

[3] BRADLEY R A. Rank analysis of incomplete block designs III. Some large-sample results on estimation and power for a method of paired comparisons[J]. Biometrika, 1955, 42(3): 450-470.

[4] ATKINSON D R, WAMPOLD B E, LOWE S M, et al. Asian American preferences for counselor characteristics; application of the Bradley-Terry-Luce model to paired comparison data[J]. Counseling Psychologist, 1998, 26(1): 101-123.

[5] MCMANUS I C, RICHARDS P, WINDER B C. Do UK medical school applicants prefer interviewing to non-interviewing schools? [J]. Advances in Health Sciences Education, 1999, 4(2): 155-165.

[6] ZEITHAML V A. Service quality, profitability, and the economic worth of customers; what we know and what we need to learn [J]. Journal of the Academy of Marketing Science, 2000, 28(1): 67-85.

[7] SINSHEIMER J S, MCKENZIE C A, KEAVNEY B, et al. SNPs and snails and puppy dogs tails; analysis of SNP haplotype data using the gamete competition model[J]. Annals of Human Genetics, 2001, 65(5): 483-490.

[8] KOEHLER K J, RIDPATH H. An application of a biased version of the Bradley-Terry-Luce model to professional basketball results[J]. Journal of Mathematical Psychology, 1982, 25(3): 187-205.

- [9] MCHALE I, MORTON A. A Bradley-Terry type model for forecasting tennis match results[J]. International Journal of Forecasting, 2011, 27(2): 619-630.
- [10] BALREIRA E C, MICELI B K, TEGTMEYER T. An oracle method to predict NFL games[J]. Journal of Quantitative Analysis in Sports, 2014, 10(2): 183-196.
- [11] BRESLOW N E, CLAYTON D G. Approximate inference in generalized linear mixed models[J]. Journal of the American Statistical Association, 1993, 88(421): 9-25.
- [12] FIRTH D, MENEZES R X. Quasi-variances[J]. Biometrika, 2004, 91(1): 65-80.
- [13] 新浪竞技风暴. 中国篮球数据库[EB/OL]. (2016-06-09)[2016-12-28]. <http://cba.sports.sina.com.cn/cba/schedule? qleagueid=171&qmonth=&qteamid=2016-06-09>.
- [14] 叶庆晖, 邓飞. 男子篮球比赛常规技术统计指标与比赛名次相关关系的分析研究[J]. 中国体育科技, 1999, 35(12): 28-29.
- [15] 张斌. 亚、欧洲男子篮球技术战术特点研究[J]. 北京体育大学学报, 2010, 33(7): 137-141.

Analysis of ability about China Basketball Association team based on Bradley-Terry model

GAO Changyu¹, LI Dan², HOU Wen², GAO Qingyu³

(1. Modern Educational Technology Center, Dalian Medical University, Dalian 116044, China;

2. School of Mathematics, Liaoning Normal University, Dalian 116029, China;

3. Confucius College Office, Dalian University of Foreign Languages, Dalian 116044, China)

Abstract: With the development of sports cause and sports science in China, data analysis and statistical works on sports have attracted more and more concerns. We analyze the data of the ability index about 20 teams and technical indicators factors affecting team's ability index within the Bradley-Terry model about China Basketball Association in 2015—2016, and the results reflect the true ability level of each team. In addition, we find that for rebound of basketball, assisting and shooting average are the main technical indicators which affect team's ability index which provide some evidence for predicting sports events and improving skills and tactics of sport teams.

Key words: paired comparison data; Bradley-Terry model; ability index