

基于模糊主成分分析的篮球比赛因素分析

Factor Analysis of Basketball Match Based on Fuzzy Principal Component Analysis

王武

Wu Wang

天津理工大学中环信息学院
中国·天津 300380

Zhonghuan Information College of Tianjin
University of Technology,
Tianjin, 300380, China

基金项目:天津理工大学中环信息学院
青年教师科研育苗基金;项目名称:模
糊数学以及序结构在数学建模中的应
用【2019KYL04】

【摘要】篮球比赛胜负因素一直是业界研究的课题,掌握篮球比赛的胜负因素及其权重对于运动员选择和训练关注的侧重点都有很大的帮助。北京首钢队作为中国篮球男子职业联赛(CBA)一支重要的队伍,近些年来一直在努力进步,有着很大提高的空间。论文选取了北京首钢队在CBA联赛2018—2019赛季的比赛中与对手的最终分差在10分以内的18场比赛的数据,应用主成分分析法,给出了影响比赛胜利的2分球命中率、篮板、助攻等9个技术统计的一个综合指标“z值”,并结合2分球命中率、篮板、助攻等9个技术统计在“z值”中的权重,讨论了北京首钢队在这18场比赛中表现出来了优势、劣势、不足,并给出了一些有意义的建议。

【Abstract】The factors of winning and losing in basketball matches have always been the subject of study in the basketball industry. As an important team in the Chinese basketball association (CBA), Beijing Shougang group has been making progress in recent years, also has great space for improvement. This paper selects the data of 19 games in which the final difference between Beijing Shougang group and its opponents in the 2018—2019 CBA season is within 10 points, using principal component analysis (pca), gives a comprehensive index "z value" of 9 technical statistics, such as 2-point field goal percentage, rebound and assist, which influence the victory of the game, and combining the two shooting, rebounds, assists, and other nine technical statistics in "z values" of weights, this paper discusses the advantages and disadvantages of Beijing Shougang group in the 18 matches, and gives some meaningful suggestions.

【关键词】CBA;北京首钢;主成分分析;z值

【Keywords】CBA; Beijing Shougang group; principal component analysis; z value

【DOI】10.36012/sde.v2i1.776

1 引言

中国男子篮球职业联赛(China Basketball Association),简称中职篮(CBA)是由中国篮球协会所主办的跨年度主客场制篮球联赛,是中国最高等级的篮球联赛^[1]。联赛自1995年以来已经发展成为中国最具观赏性、竞技性的体育联赛,CBA目前有20支球队,每个赛季共计460场比赛。

竞技体育的最大魅力就是比赛的不确定性,在篮球比赛中,比赛的胜利是每只CBA球队追求的目标^[2-4]。但是影响比赛胜利的因素有很多,主要包括2分球命中率、3分球命中率、罚球命中率、篮板球、助攻、盖帽、抢断、失误、犯规等因素。本文以CBA联赛北京首钢队为例,选取北京首钢队在2018—2019赛季与对手比赛的最终分差在10分以内的18场比赛。

主成分分析是化多元影响因子为少数综合指标,并给出

对应权重的一种统计方法。本文将采用主成分分析法,将影响比赛胜负的9个主要因素进行整合,并得到整合后的影响因子,同时找出上述9个主要影响因素的权重,从而分析上述9个因素对北京首钢队比赛的影响情况,并给出一些建议。

2 数据标准化

未处理的数据具有离散性、随机性等特点,故需先对数据进行处理,具体步骤为:

①在上述9个影响因素中,犯规和失误为负面影响因素,故对这两列数据取相反数。

②将①中处理后的数据记为矩阵 $X=(X_{ij}), i=1, \dots, 18, j=1, \dots, 9$ 即用 $X_i, i=1, \dots, 20$ 分别表示相应轮次比赛的数据,如 X_5 表示第9轮比赛, $X_5=(X_{51}, \dots, X_{59})$ 表示第9轮比赛的各项数据。

记第 i 列数据的平均值 $\bar{X}_j = \frac{1}{18} \sum_{i=1}^{18} X_{ij}$, 标准差为 $\sigma = \frac{1}{18} \sum_{i=1}^{18} (X_{ij} - \bar{X}_j)^2$, 则第 j 列数据的标准化为 $\bar{X}_{ij} = \frac{X_{ij} - \bar{X}_j}{\sigma}$ 。对选取的比赛数据进行标准化操作, 可以得到标准化数据。

3 构造协方差矩阵

在统计学与概率论中, 协方差矩阵的每个元素是各个向量元素之间的协方差, 是从标量随机变量到高维度随机向量的自然推广。

令 $c_{ij} = Cov(X_i, X_j)$ 表示 X_i 与 X_j 的方差, 则

$$C = (c_{ij})_{n \times n} = \frac{1}{n-1} \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

容易计算, 上述的矩阵 c 即为协方差矩阵, 计算得:

$$C = \begin{bmatrix} 0.125 & -0.014 & -0.003 & 0.004 & -0.001 & -0.050 & -0.036 & -0.022 & -0.012 \\ -0.014 & 0.126 & 0.013 & 0.000 & 0.064 & 0.022 & 0.020 & -0.001 & 0.029 \\ -0.003 & 0.013 & 0.125 & -0.047 & 0.028 & 0.036 & -0.005 & 0.048 & -0.007 \\ 0.004 & 0.000 & -0.047 & 0.125 & -0.008 & -0.001 & -0.040 & -0.090 & -0.019 \\ -0.001 & 0.064 & 0.028 & -0.008 & 0.125 & -0.002 & 0.008 & 0.025 & 0.028 \\ -0.050 & 0.022 & 0.036 & -0.001 & -0.002 & 0.125 & -0.029 & 0.011 & 0.019 \\ -0.036 & 0.020 & -0.005 & -0.040 & 0.008 & -0.029 & 0.125 & 0.057 & -0.004 \\ -0.022 & -0.001 & 0.048 & -0.090 & 0.025 & 0.011 & 0.057 & 0.125 & 0.015 \\ -0.012 & 0.029 & -0.007 & -0.019 & 0.028 & 0.019 & -0.004 & 0.015 & 0.125 \end{bmatrix}$$

4 主成分分析

通过 matlab 计算得协方差矩阵 R 的 2 个最大特征值为 0.4、1.3:

相应的第一、第二主成分为:

$$F_1 = (x_1, \dots, x_9) a_1 = -0.03x_1 + 0.23x_2 + 0.36x_3 - 0.48x_4 + 0.28x_5 + 0.17x_6 + 0.32x_7 + 0.38x_8 + 0.19x_9$$

$$F_2 = (x_1, \dots, x_9) a_2 = 0.33x_1 + 0.51x_2 + 0.02x_3 - 0.17x_4 + 0.38x_5 + 0.40x_6 - 0.29x_7 - 0.28x_8 + 0.33x_9$$

显然 1.246 与 0.408 两个特征值之和为 1.654, 所有特征值和为 2.271, 则这两个特征值为特征值之和的 73% > 70%。则选取数值最高的特征值对应的特征向量作为新的指标。

指标与为整合之后的新的影响比赛的因素, 这两个指标的累计贡献率为 73%, 故其在很大程度上反映了影响比赛的各种因素的权重情况。现只考虑 F_1 、 F_2 两个指标, 指标所占权重为 24%, 指标所占权重为 76%。根据两个特征向量比重, 故构造新的综合评价指标为:

$$z = 0.10x_1 + 0.30x_2 + 0.27x_3 - 0.41x_4 + 0.31x_5 + 0.22x_6 + 0.17x_7 + 0.21x_8 + 0.22x_9$$

相应场次的 z 值如表 1 所示:

表 1 相应 z 值

1	4	5	6	9	10	17	18	20
-0.16	0.12	0.13	-0.33	-0.18	-0.06	-0.24	0.21	-0.01
24	30	31	33	36	39	40	43	45
-0.11	-0.14	0.58	-0.08	-0.03	-0.49	0.27	0.47	-0.02

z 值越高, 说明北京首钢队在本场比赛中整体发挥越好, 取得比赛胜利的概率越大。

5 结果分析与对北京首钢队引进外援、日常训练的建议

通过对北京首钢队 18 场比赛的分析整合出了一个评判球队竞技水平的一个综合指标 z 值, 由此可以得到一些结论:

①在 z 值中, x_4 的系数为 -0.41, 说明在北京首钢队的这 18 场比赛中, 篮板球对比赛的胜利有负作用, 即北京首钢队在这两方面的竞技水平低于对手。通过分析发现, 北京首钢队在上述 18 场比赛平均篮板球数为 34.89, 场均低于对手 7.3 个, 北京首钢队在 2018—2019 赛季 46 场比赛中, 场均篮板球数为 36.1, 为全联赛最低。这也说明, 篮板球已成为北京首钢队最大的短板。

②在 z 值中, x_5 的系数为 0.31, 这说明北京首钢队在这 18 场比赛中, 助攻对比赛胜利的贡献最大, 对球队战术执行较好。通过分析发现, 北京首钢队在上述 18 场比赛平均篮板球数为 19.39, 场均高于对手 2.1 个, 北京首钢队在 2018—2019 赛季 46 场比赛中, 场均篮板球数为 20.8, 为全联赛第二。

③在第 39 轮对阵广州队的比赛中, 北京首钢队的 z 值最小, 说明球队在这一轮发挥较差, 命中率、助攻等数据均低于平局水平, 同时得分为 93 分, 低于赛季平均分近 8 分, 最终以比分 93:97 输给了广州队。

④在第 31 轮对阵上海队的比赛中, 北京首钢队的 z 值最大, 说明球队在这一场比赛中发挥较好, 场均得分、助攻、抢断均高于平均水平, 尤其是助攻为 35 次, 为首钢队赛季最高, 最终以比分 120:118 战胜上海队。

参考文献

[1] 薛玉行. 2009-2010 赛季 CBA 常规赛天津荣钢队主场进攻技术数据统计分析[J]. 运动, 2011, 30(2): 16-18.
[2] 闫玉东, 关凯, 赵晶. CBA 职业联赛京津主场球市开发现状及策略[J]. 体育成人教育学报, 2018, 34(4): 55-57.
[3] 胡诗韵, 张焱. 基于主成分分析的江苏省人口城镇化质量评价[J]. 中国物价, 2018, 34(3): 35-37.
[4] 赵丽娟, 董小林, 吴阳. 基于 Z-Score 模型的环境污染与经济发展的联合分析[J]. 应用化工, 2017, 46(9): 1805-1809.