

2012-13 赛季英超球员数据统计分析

一、研究问题

英格兰足球超级联赛 (Premier League)，通常简称“英超”，是英格兰联赛系统的最高等级联赛。英超是“欧洲五大联赛”之一，由 20 支球队组成，英超联赛一直以来被认为是世界上最好的联赛之一，节奏快、竞争激烈、强队众多，现已成为世界上最受欢迎的体育赛事，也是收入最高的足球联赛。随着科技以及数据产业的发展，应用统计学知识来合理的量化球员们在球场上的表现，评价球员价值以及指导球队完善人员体系成为了足球世界中不可或缺的一部分。同时，利用数据来挖掘球队与球员的潜力也成为了推动现代足球发展的强劲主力。本分析报告利用被记录的来自 16 支球队的部分球员（166 位）的各项指标来进行以下问题的讨论：

- 1、英超联赛的总体宏观统计分析
- 2、各位置球员进球能力的评估
- 3、对曼彻斯特联队的球员体系进行详细分析
- 4、球员数据各项指标的内在关联分析

二、变量介绍

首先导入数据，然后使用 `str`, `summary` 命令可以大致了解到数据包含 166 个样本，34 个变量。可以做出下表更加详细的列出变量信息。

表 2-1 英超球员统计变量表

变量类型		变量名	详细说明	备注
因变量		下赛季进球数	定量变量 单位：个	
自变量	基本状况	球员姓名、球员年龄、所属球队、球衣号码、场上位置等	场上位置为定性变量 有前锋、中场、后卫三个水平	
	控球能力	传球次数、过人次数	定量变量 单位：次	
	进攻能力	助攻、射门次数、进球射正次数、赢得点球	定量变量 单位：次	进球数包含头球、左脚、右脚、任意球、点球、乌龙球等
	防守能力	拦截、解围	定量变量 单位：次	解围包括头球解围和后场解围
	出场能力	出场次数、首发次数、出场时间	正相关性明显 定量变量	出场时间单位由分钟转化为小时
	推进能力	抢断、头球争顶成功	定量变量 单位：次	
	犯规状况	越位、犯规、红牌、黄牌	定量变量 单位：次	两张黄牌累计红牌，犯规为未得牌的犯规

三、描述分析及结论

1、英超联赛的总体宏观统计分析

下图中统计出了 166 名球员中的个人进球个数的统计频数，可以看到整体成右偏分布，大多数球员进球数量为 0-5 个，共有大约 128 人。有少数天才进攻性球员进球数相对来说很多，达到 15-30 个。其中 2012 到 2013 赛季的最佳射手前三名分别是范佩西、苏亚雷斯和贝尔，进球数分别是 26 个、23 个和 20 个。正好对应着直方图的尾部的三个统计分组。天才进攻型球员的发挥往往决定着一个球队的进攻水平。

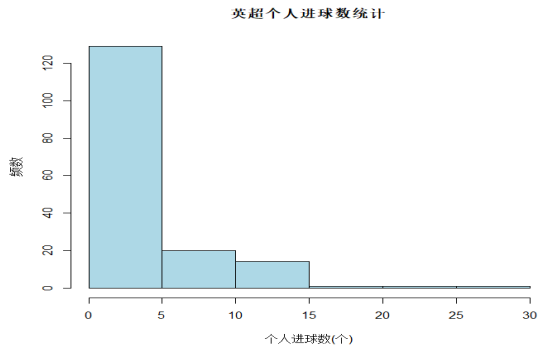


图 3-1 英超个人进球数直方图

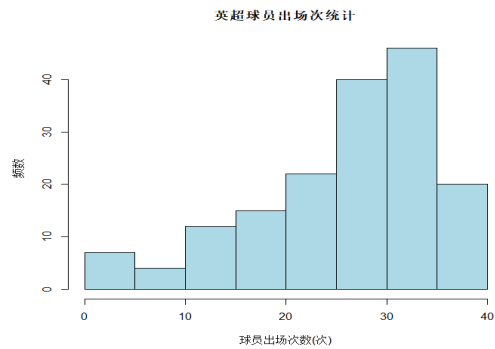


图 3-2 英超球员出场次数直方图

由图 3-2 可以看出英超球员的整体出场次数大概成左偏分布，这意味着大多数球员都拥有着相对较高频率的出场次数这有利于维持球员的竞技状态，但也存在着少数边缘球员极少的获得的上场机会较少，在 38 轮的联赛中仅获得 0-10 次的出场机会，这不利于维持竞技水平，也预示这该名球员在球队中国的重要性较低，下赛季极有可能离队。

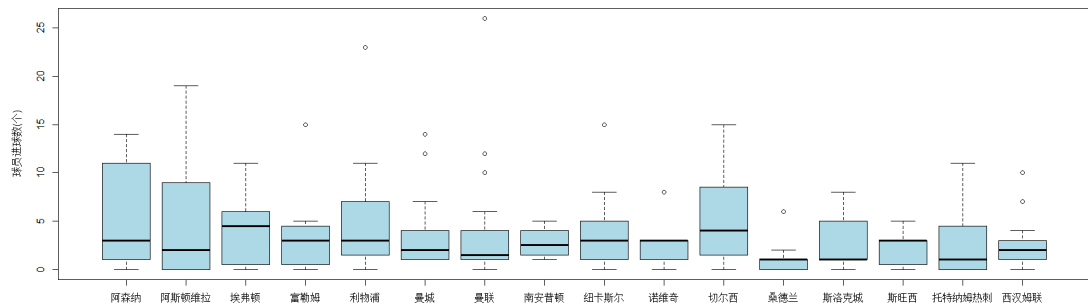


图 3-3 各球队球员进球数分布

由图 3-3 可以看出箱线图中箱子比较长的说明球队队员的进球能力差距较大，扁箱子意味着球员们的进攻能力相差不多，可以看到像切尔西，阿森纳等球队的进球能力差距较大。而各个箱线图的箱外离散点可能是该队的前锋，他们拥有着强大的进攻火力。像曼联队的范佩西，进 26 球，其他球员进球能力相似，所以这样的球队阵型明确，分工合理，前锋强大。

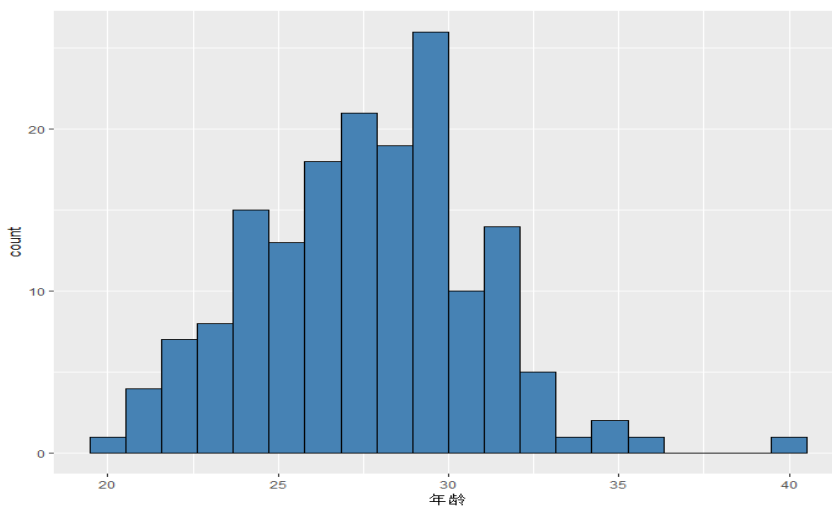


图 3-4 英超各年龄段人数统计直方图

一个联赛的年龄构成也反映了该联赛的运营状况。一个合理而又健康的联赛一定是既拥有着年轻的血液又有老当益壮的老将。从图 3-4 中我们可以看出英超联赛的运营很健康，年龄结构呈钟形分布，新老交替完美大多数球员的年龄在 24-32 岁左右，这也是一个球员的职业生涯的黄金期。但也有像费迪南德和吉格斯这样的老将活跃在赛场上。

2、各位置球员进球能力的评估

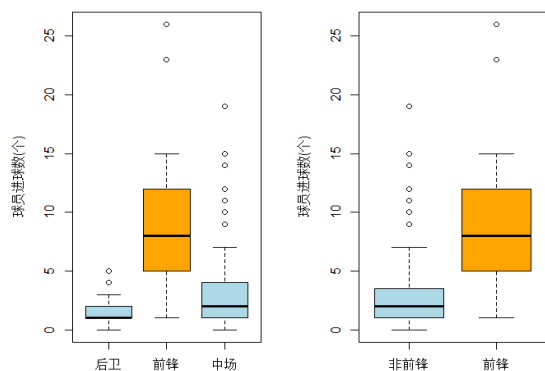


图 3-5 不同场上位置对进球的影响

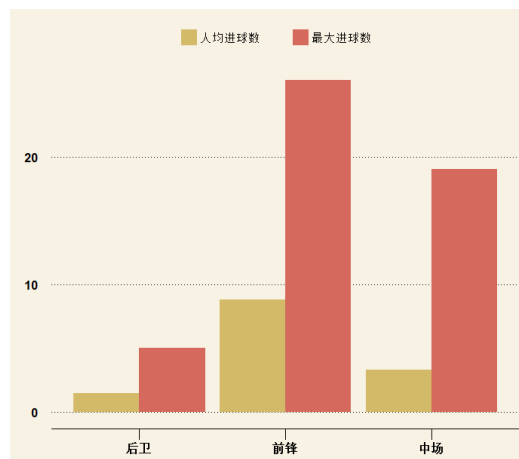


图 3-6 不同场上位置的进球统计

足球发展到现在讲究体系配合和团队足球，即每一个足球队场上的球员各司其职，形成一个体系，整体的进攻与防守，这也是源于越位规则的完善。阵型足球使得每个球员的特点被发挥出来，更具有审美性。那么这样明确过后我们应该明白前锋的射门机会将远超其他位置的球员，而且门前嗅觉灵敏自然进球较多，从图 3-5 和图 3-6 上可以看出离球门更近前锋进球平均数要远高于其他位置的球员最多进球 26 个，平均进球也有 9 个，但箱线图中可看出前锋们的进球差距较大，这不仅与前锋水平有关，还与中场组织和后卫插上有关。其次中场球员的偶尔前插助攻，组织进攻也带来进球仅次于前锋进球量。最后是主要负责防守和拦截的后卫除了定位球的机会在运动战中很少斩获进球，平均进球数量只有 2 个，最多的进球量也就有 5 个，分布较为均匀。

3、对曼彻斯特联队的球员体系进行详细分析

曼彻斯特联队作为英超老牌豪门，几乎获得了足球俱乐部能获得的所有荣誉，2012-13 赛季更是获得了英超联赛冠军，接下来通过雷达图来综合分析曼联不同位置的球员，来分析体系结构，战术安排以及人员的轮换。我通过表 2-1 英超球员统计变量表的能力变量的所属变量除以出场时间转化为效率，然后离差标准化，在加和取平均来形成综合能力。详细过程见 R 文件。

表 3-1 中场综合能力

Name	Trap	Attack	Defense	Importance	Advance
安德森	0.534283	0.15425	0.072321	0.583819	0.165299
吉格斯	0.364023	0.20821	0.070838	0.22474	0.215766
埃尔南德	0.179438	0.694151	0.043884	0.320549	0.06167
卡里克	0.525096	0.096514	0.203073	0.319416	0.169836
纳尼	0.322705	0.300874	0.105353	0.376223	0.108158
阿什利杨	0.297002	0.198105	0.090786	0.27678	0.119877
克莱弗利	0.426897	0.148729	0.108646	0.277478	0.146195
香川真司	0.417233	0.310656	0.051395	0.140156	0.083366
巴伦西亚	0.433376	0.143043	0.088551	0.212184	0.195609

形成雷达图，如图 3-7 中场球员的能力分析发现，作为中场球员，组织进攻与防守，所以推进和控球能力比较重要，也可以看出曼联的中场球员都具有相似的推进和防守能力，其中安德森的推进和控球能力突出且上场时间长，而“小豌豆”埃尔南德斯的进攻能力较强，可以在不同情况下调用各名球员到不同位置。

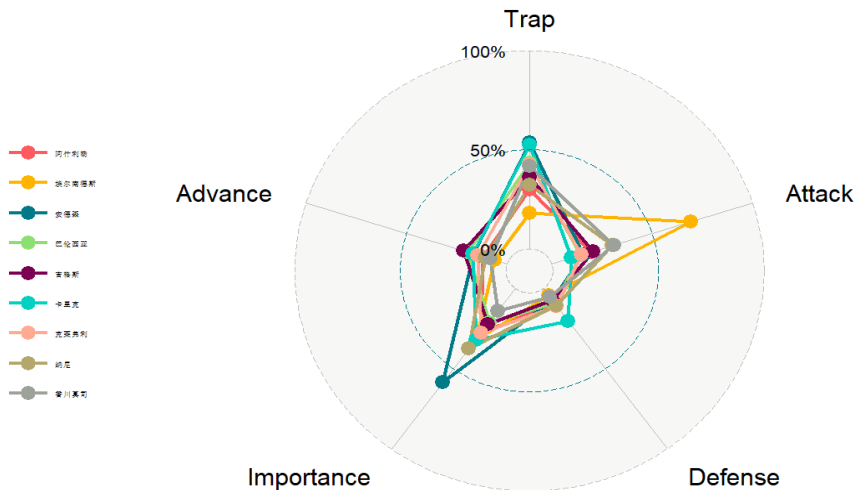


图 3-7 中场球员能力分析图

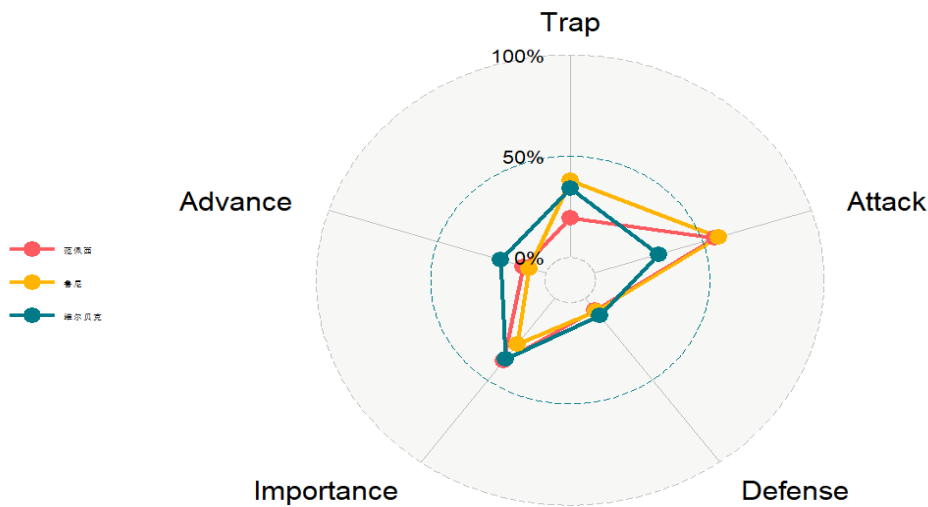


图 3-8 前场球员能力分析图

如雷达图 3-8 所示很明显前场球员能力主要集中在进攻和控球上，作为传统高又硬的站柱中锋的鲁尼，拥有强悍的身体素质所以作为球队在前场的支点，也具有良好的控球能力。而如图 3-8 来看也恰好反映了他们的特点具有良好的防守能力，其中维迪奇的防守能力最强，而斯马林由于他在 2013-13 的不佳表现也被淘汰除了球队。

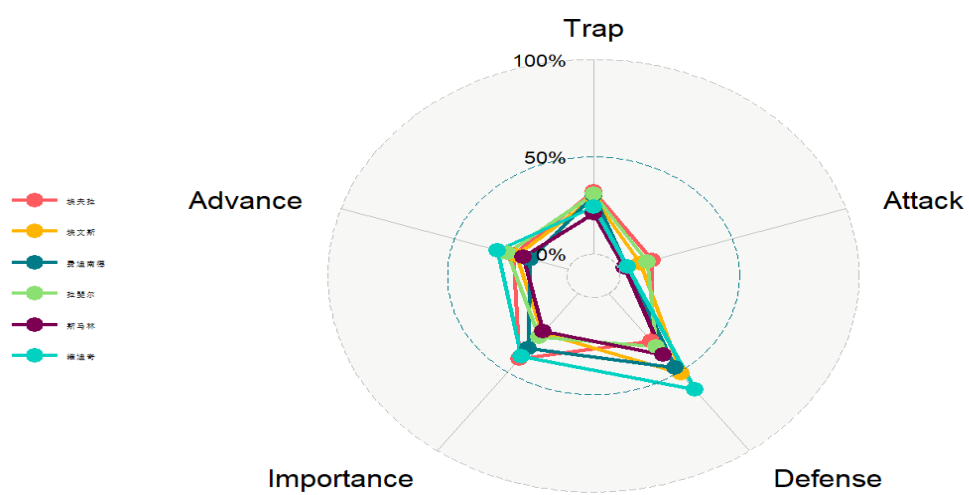


图 3-9 后场球员能力分析图

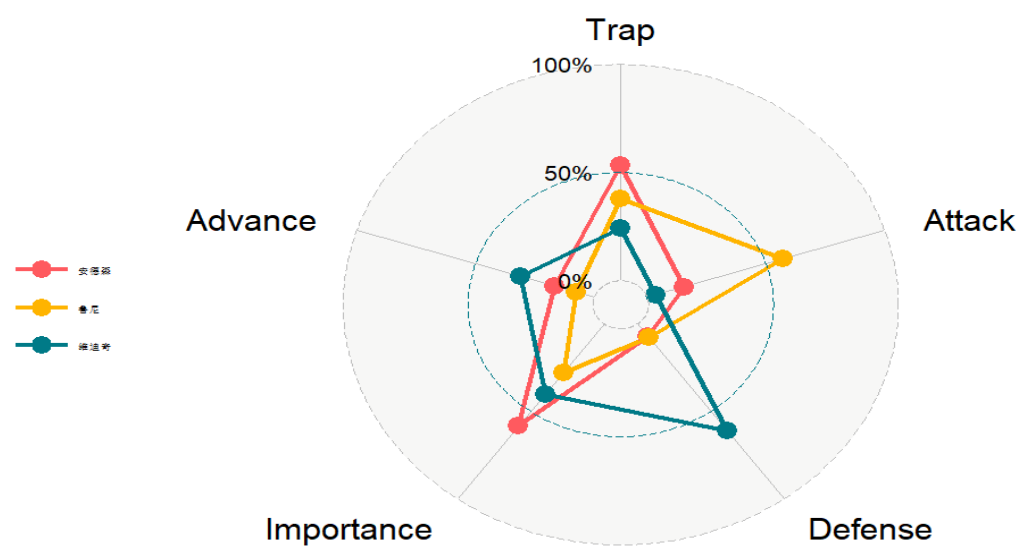


图 3-10 曼联不同位置球员能力分析图

图 3-10 来看选取了曼联的几个位置上的典型球员，也突显了他们位置上的各自特点，后卫—防守，中场—控制，前锋—进攻。

4、球员数据各项指标的内在关联分析

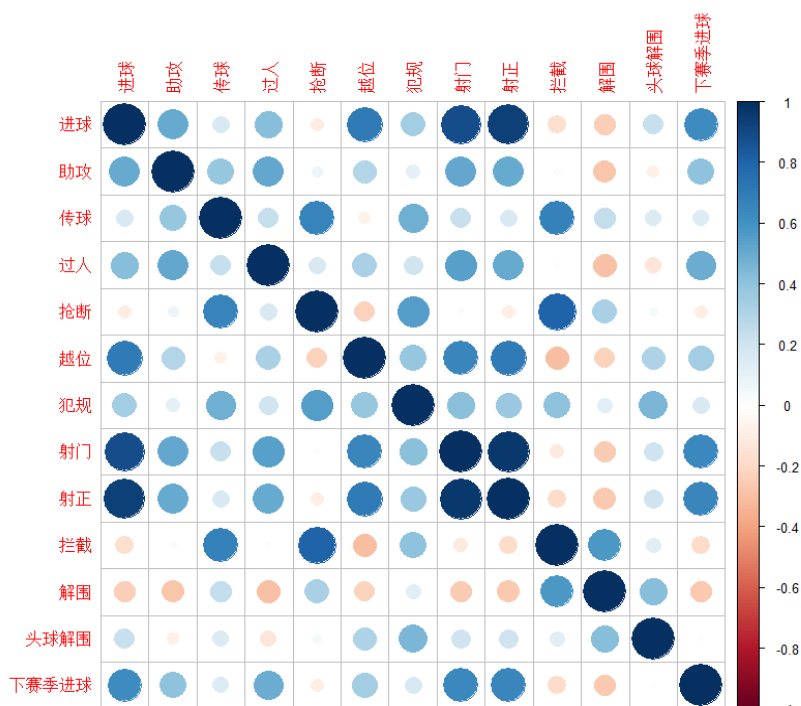
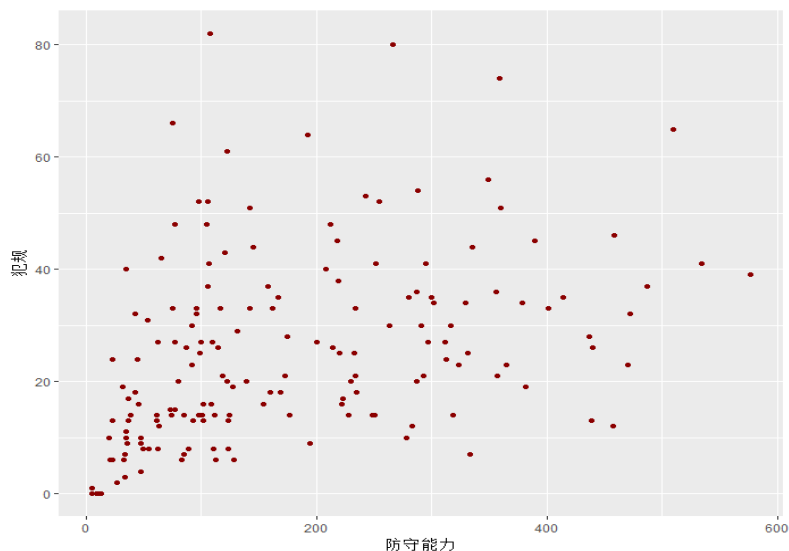


图 3-11 各项指标之间的相关关系

如上图所示的球员个指标的相关关系，红色点为负相关关系，蓝色点为正相关关系，点越大则相关关系越大，可以看出有很多指标拥有极强的共线性，因此建模时应应用减轻共线性方法来消除对建模的影响。

我们感兴趣的是防守容易造成犯规，那是否一个球员的防守能力越强他的犯规次数就会越强呢，因此分析综合出的防守能力与球员的犯规次数之间的关系，如图 3-12。



如图所示，两者之间的关系在防守能力小的时候比较呈正相关，但随着数值的增大相关关系并不明显，因此可以说明一个球员的防守积极性并不与犯规次数成正相关，还是更考虑球员的个人能力。