

第一章：绪论：

1.1 研究背景：

不久之前，金州勇士队通过分析大量的 NBA 赛事数据和观看大量的球队比赛，找到了防守安东尼·戴维斯的最佳球员。在最新的一赛季比赛中，勇士队的表现有了新的飞跃和突破。今年，几乎 NBA 每一支球队都开始追踪其球员的比赛数据，包括运动员在场上的得分位置，或者每个队员的比赛数据等。通过有效的数据分析，高级的数据模型，和分析工具，NBA 比赛已经逐渐转化为专业篮球比赛。这样的转变不仅影响着球员的打球习惯，教练的训练方法，粉丝与球星的互动方式，甚至在赛制上也有了不同的调整运动员，教练，粉丝，甚至 NBA 分析员都希望好好利用大量的数据，来满足他们不同的需求。篮球比赛在每个常规赛季中进行 82 场比赛。网上公开了大量的比赛数据，也有很多写得很好的文章和书籍分析球队球员表现。但是专业分析员会利用每场比赛球员的场均走位，以及更高级的指标进行更具体详细的分析。

近几年，衡量球员表现的指标数据在近几年间不断更新，更加全面的反映球员的表现。最基本且最容易获得的数据是每个球员和球队每场比赛的得分数据。得分数据不仅记录某个球员和他队友的各项得分，还记录了他的对手球队和球员的防守数据。之前，得分数据仅仅包含最基础的球员数据，包括球员进球得分，篮板球，助攻数，和投篮命中率等。但这些数据来衡量一个球员的表现，能力和价值是远远不够的。首先，现有的统计量更多偏向记录球员的进攻数据，而忽视了球员的防守贡献水平；如果这个球员在整场比赛中投篮次数更多，助攻更多，这个球员可能是更有价值的球员，但尽管他有很强的防守能力，比如他多次抢断对方投球，或盖帽，他也不一定能有很高的综合评分。其次，现有的统计数据可能对在场上拥有更多控球权的球员更有利；比如控球后卫将球带到前场并且更多的组织进攻，他带球时间更长，所以他的综合数据可能就更优秀。但是一个球员的控球能力却没有很精准的数据衡量，但控球能力是一个很重要的衡量指标。因此联盟从 1970-1971 年的赛季之后，开始记录球员的犯规数。

至今，由于有了更精准的记录仪器，我们可以获得更准确地球员数据，进行更具体的数据分析。例如金州勇士队，就大量利用大数据分析对球队整体训练计划和决策等方面进行改正。本文不会用到每一个球员在场上的实时位置信息和训练信息等十分具体的数据进行预测，因为这些数据不是对大众公开的。但是本文将会运用更常见的统计指标如射门得分，助攻次数，抢断次数，盖帽次数等，通过分析这些指标对球员和球队贡献的重要性来预测该球队的输赢。

1.2 问题提出：

本文以 NBA 官网公布的数据和信息为基础，探究如何更综合得分析球员实力，球员自身表现如何影响球队整体表现，教练如何制定更适合某队训练计划和人员排布，以及观众和一些商业运作人员如何根据基础数据预测出下一赛季的比赛结果。

1.3 研究意义：

理论意义：本研究试图构建全面衡量球员表现的高级指标，综合成可以衡量某队球员整体表现的球队指标，与衡量球队表现的高级指标，通过建立统计学模型，寻找影响球队胜率的主要原因。

实际意义：通过理论分析，研究人员记录和跟踪运动员的身体数据可以制定出更合适的训练方法，对十分有价值的球员可以针对他设计出一个更能够辅助他得到更多分数的球队阵容从而为球队赢得更多的比赛；在娱乐领域，观众可以根据 NBA 官网公布的球

员球队数据和信息，计算和预测球队的胜负。本文希望根据个体球员在比赛中的的基本数据得到反映其对所在球队贡献程度的指标，即衡量球员价值的指标，综合某个球队所有球员的贡献，得到球队的整体实力，从而根据球队的整体数据对下一个赛季中球队的胜率进行预测。

1.4 论文框架结构：

第一部分首先进行文献综述，总结并介绍国内外相关的研究以及本文所用到的参考文献；第二部分总结网上公开的 NBA 数据名称，数据在本文中的用途及其背后的含义，根据可以找到的基础数据构建高级篮球分析指标并进行数据的描述性分析，更全面的认识本文构建的高级篮球分析指标；第三部分首先进行指标间的相关性分析，为建立模型做铺垫，根据相关性分析排除相关性极高的变量后，根据模型的指标建立多元线性回归模型，根据该模型的结果进行模型的调整，使用三种不同的调整方法，包括岭回归，主成分回归和只用方差膨胀因子小的变量在此建立多元线性回归模型。最终得到最佳的统计模型，该模型的显著性较高，更好的解决了本文提出的问题。

第二章：文献综述

2.1 国外研究

在国外 Chatterjee, Campbell 和 Wiseman(1997,)对一个赛季的数据建立统计模型，对球队胜率进行回归分析，发现比赛得分、罚球、篮板球和失误在统计上是显著的，并且回归系数在隔年数据之间都是相对稳定的；Hausman 和 Leonard(1997,)使用机灵经济学的方法对 NBA 赛事明星出场率与电视收视率和门票收入进行相关性研究，得出正向结论；Mizak, Stair 和 Rossi(2004)使用胜率标准差、HHI 等指标衡量了各大联盟的竞争性平衡，说明了所使用指标的优缺点，并指出：指标的缺陷不是最重要的，最重要的是用同一指标观察联盟竞争性平衡的趋势；Kubatko, Oliver, Rosenbaum(2007)首次介绍了高级篮球分析指标的相关概念以及在研究领域无法统一的公式上进行了比较与总结；Franks, Miller, Bornn, Goldsberry 等人构建了高级防守指标，结合运动员在比赛中的首发阵容和走位，引入了一套新的防守指标，对球员跟踪、结合统计建模和可视化分析能够比以前更丰富地描述防守。该方法结合更传统的进攻统计，提供了一个全面的总结，球员的贡献，以一场比赛的最终结果；在预测两只球队未来比赛结果方面，Hu 和 Zidek(2002)使用相关加权似然法(i)使用所有相关的样本信息；(二)体现主场优势的关联概率法进行预测。

2.2 国内研究

由于 NBA 进入中国市场的时间不长，国内的研究也不太丰富。通过对国内 NBA 比赛问题研究的相关文献的整理和回顾，国内有关 NBA 的赛事研究出了借鉴国外成国外，其研究特色分为：（1）从市场营销的角度对 NBA 的市场价值、品牌文化传播何在中国的市场营销等方面对 NBA 的品牌建设进行剖析；（2）从制度经济学角度对 NBA 的人力制衡，收益制衡及权力制衡三项机制的功能及相关值得运行原理展开深入研究。（3）NBA 比赛本身的技术角度，如球赛中冲抢技术、不同位置球员攻受能力、求赛赛程安排等角度进行研究分析。

2.3 小结

根据以往文献，众多统计学家试图从众多影响因素中找到影响球队胜率的最显著因素。但由于缺少对高级指标的研究，只根据现有的基础数据无法全面的构建球员

的技术能力。本文通过构建高级指标更综合与全面的界定球员表现与技术，并融合球员数据到球队整体数据，从微观和宏观的角度分析，不仅给了教练在制定运动员个体训练计划上有理论支持，还更科学的规范了球队整体的训练计划和在比赛中的阵容搭配。