本文采用的是近10个常规赛季的公开数据进行建模。由于比赛从赛制，到规则等各个方面都随着时代不断改变，甚至评判输赢的规则都有改变；尽管目前可以收集到自从1946-47赛季开始至今的所有赛事数据，之前的数据对现在比赛的分析也没有太多可借鉴作用。所以我们用今年来的所有常规赛事所公开的数据进行分析与建模。比如在上世纪末期，一个球队的输赢很大程度取决于个头最大的球员（如中风和大前锋），因为当时比赛的节奏相对较慢，主导比赛进程的球员通常是个头大的球员；但是如今，比赛节奏越来越快，比赛进程也更趋于数据化，专业化和技术化，这就为身形较小而灵活，但是技术水平很高，如投篮水平很高，控球能力很强的球员如史蒂芬库里，詹姆斯哈登，凯利欧文等身高不出众但命中率很高的球员。但联盟赛制转变后更有利于进攻型球员，尤其是擅长投射的球员，这使得比赛更激动人心，吸引更多观众的目光。

* 每个赛季所有球员的整体数据，变量包括

以2011-12赛季比赛中球员数据为例：

|  |  |  |
| --- | --- | --- |
| 变量名 | 变量含义（英文） | 变量含义（中文） |
| Rk | Rank | 排名 |
| Pos | Position | 首发位置 |
| Age | Player's age | 年龄 |
| Tm | Team | 所在球队 |
| G | Games | 该赛季比赛总场数 |
| GS | Games Started | \*\*\* |
| MP | Minutes Played | 球员上场时长 |
| FG | Field Goals | 球员投篮命中 |
| FGA | Field Goals Attempts | 球员投篮 |
| FG% | Field Goal Percentage | 场均球员命中率 |
| 3P | 3-Point Field Goals | 场均3分球命中得分次数 |
| 3PA | 3-Point Field Goal Attempts | 场均3分球投篮次数 |
| 3P% | FG% on 3-Pt FGAs | 场均3分球得分率 |
| 2P | 2-Point Field Goals | 场均2分球命中得分次数 |
| 2PA | 2-point Field Goal Attempts | 场均2分球投篮次数 |
| 2P% | FG% on 2-Pt FGAs. | 场均2分球得分率 |
| eFG% | Effective Field Goal Percent | 场均有效的投篮得分率 |
| FT | Free Throws | 场均罚球得分 |
| FTA | Free Throw Attempts | 场均罚球投篮次数 |
| FT% | Free Throw Percentage | 场均罚球命中率 |
| ORB | Offensive Rebound | 场均进攻篮板球次数 |
| DRB | Defensive Rebounds | 场均防守篮板球次数 |
| TRB | Total Rebounds | 场均篮板球总次数 |
| AST | Assists | 场均助攻次数 |
| STL | Steals | 场均盖帽次数 |
| BLK | Blocks | 场均抢断次数 |
| TOV | Turnovers | 场均失误次数 |
| PF | Personal Fouls | 场均个人犯规次数 |
| PTS | Points | 场均得分 |

* 收集每个赛季所有球队的信息如下：

以2011-12赛季比赛中球员数据为例：

|  |  |  |
| --- | --- | --- |
| 变量名 | 英文名称 | 中文解释 |
| Rk | Rank | 球队在赛季中的排名 |
| W | Wins | 球队整赛季赢过的比赛 |
| W/L% | Win-Loss Percentage | 球队的胜负率 |
| MOV | Margin of Victory | 输赢球队比分之差 |
| Ortg | Offensive Rating | 每100次进攻的得分 |
| DRtg | Defensive Rating | 每100次进攻的失分 |
| NRtg | Net Rating | ·每100次进攻机会的净胜分 |
| MOV/A | Adjusted Margin of Victory | 根据对手进攻节奏调整后的MOV |
| Ortg/A | Adjusted Offensive Rating | 根据对手进攻节奏调整后的每100次进攻得分 |
| DRtg/A | Adjusted Defensive Rating | 根据对手进攻节奏调整后的失分 |
| NRtg/A | Adjusted Net Rating | 根据对手进攻节奏调整后的净胜分 |

* 收集每个球队的整个赛季中与之对抗的球队的比赛数据，变量名如表一。
* 每个球队在整个赛季中数据统计，变量名如表一。