

NBA 最高勝率チームの統計的分析

2007MI053 本田 修大

指導教員：木村 美善

1 はじめに

バスケットボールは確率のスポーツである。シュートが外れる時は手前に約 30%、奥に約 70%の確率で落ちると統計的に出されていたり、多くの部分で確率統計的要素のあるスポーツである。このようなスポーツで、強いチームには、どのような試合が多いか、ホームとアウェーでは、どのような違いがあるかについて分析することを、卒業研究とした。

2 データについて

NBA の 07-08,08-09,09-10 年レギュラーシーズンの最高勝率チームを対象とし、シーズン 82 試合をホーム・アウェーに分けて分析した ([2] 参照)。

2.1 用いた変数

「FGA:2P,3P を打った数」、「FGM:2P,3P の入った数」、「3PA:3P を打った数」、「3PM:3P の入った数」、「FTA:フリースローを打った数」、「FTM:フリースローの入った数」、「REB:シュートミスを保持した数」、「STL:相手が保持していたボールを奪った数」、「TO:ミスした数」、「PTS:一試合の得点」

3 分析方法について

主成分分析、クラスター分析および、重回帰分析を行った。

4 主成分分析 ホームとアウェー

ホームとアウェー両方とも、第五主成分までを用いて分析を行った。第一主成分ではシュートに関する特徴、第三主成分ではミス、ディフェンス (DF) に関する特徴が現れ、ホーム、アウェーともに同様の特徴が見られる。しかし、第二主成分では、ホームが試合展開の速さと DF、アウェーがミス、第四主成分では、ホームがオフェンス (OF)、DF どちらに重点を置いているか、アウェーが OF の戦略、第五主成分では、ホームが OF の戦略、アウェーが試合の戦略となっており、ホームとアウェーで特徴が違っている。この違いから、ホームでは DF に重点を置いており、アウェーでは OF に重点を置いていると考えられる ([1],[2],[3] 参照)。

5 クラスター分析 ホームとアウェー

5.1 試合結果からのクラスター分析

ワード法を用いてクラスター分析を行った。ホームでは OF のいい試合は無敗で試合数が多く、DF のいい試合も無敗で試合数が少ない。また、OF の悪い試合、DF の悪い試合の群があるが、OF の悪い試合の群の方が負け試

合が多くなっている。アウェーでは、OF のいい試合の群は無敗だが試合数が少なく、DF のいい試合の群は負け試合がある。DF の悪い試合の群を見ると、他のどの群と比べてみても、あきらかに負け試合の数が多くなっている。よって、クラスター分析の結果から、ホームでは OF、アウェーでは DF が良ければ勝ちに繋がりやすいと考えられる ([2],[3] 参照)。

5.2 主成分得点からのクラスター分析

試合結果のクラスター分析とは、異なる結果が得られるが、主成分得点からのクラスター分析の方が試合内容について詳しく表せているので、こちらの方がいいと言える。ホームでは、ミスの多い試合で負ける可能性が高く、OF は良いが DF の良くない試合で負け試合があることから、勝敗の理由として、OF より、DF の方が影響が大きいと考えられる。アウェーではシュートが入らずに負けている試合が多いため、DF より、OF の方が影響が大きいと考えられる。

また、群ごとの試合数を見ると、ホームでは FG 成功率の高い試合が多く、アウェーでは少ない。これは、他の年のクラスター分析の結果を見ても言えることで、ホームの方がシュートが良く入ると言える。理由として、会場の違いにより雰囲気やゴールの距離感が違って来るため、年間 41 試合行うホームでは感覚がつかめているが、年 1,2 試合しか行わないアウェーではそれがかみにくいと考えられる。

また、試合結果と主成分得点でのクラスター分析の違いとして、試合結果の方では群分けの要因として OF の要素が強いが、主成分得点の方では群分けの要因として、DF について強くなっていることが挙げられる。

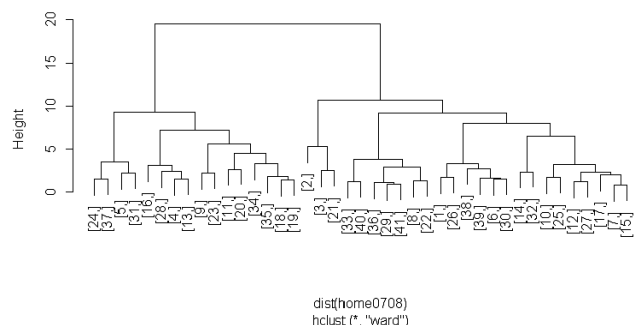


図 1 主成分得点、ホーム：デンドログラム

6 重回帰分析

主成分得点からのクラスター分析の結果のように、本当にホームとアウェーで得点に違いがあるのかを見るため、ホームとアウェーを分けた場合と一緒にした場合で重回帰分析を行った。ホームとアウェーを分けた場合は、PTS を従属変数とし、FGA,FGM,3PA,3PM,FTA,FTM,REB,STL,TO を説明変数とした ([1],[3] 参照)。

6.1 分析結果 ホーム

表 1 重回帰分析結果 ホーム

	回帰係数	標準化残差	t-統計量	p-値
切片	-97.505	3.5428	-27.522	2×10^{-16}
FGA	0.9434	0.0291	32.384	2×10^{-16}
FGM	0.1492	0.0033	44.940	2×10^{-16}
3PA	0.3389	0.0375	9.020	4.77×10^{-10}
3PM	0.0179	0.0016	10.924	5.62×10^{-12}
FTA	0.7735	0.0248	31.079	2×10^{-16}
FTM	0.0261	0.0018	14.060	9.67×10^{-15}
STL	0.1116	0.0530	2.103	0.044

総当たり法を用いて変数選択を行ったところ、FGA,FGM,3PA,3PM,FTA,FTM,STL が説明変数となった。外れ値として出た 3,4,38 試合目を取り除き、重回帰分析を行った。得点に直接関係している変数以外で、STL が説明変数になっている事から、ホームでは DF でボールを奪って、得点に繋がっていると言える。外れ値である試合を見ると、3 試合目は OF が非常に良く、4 試合目は DF の良さが OF に繋がっていない、38 試合目は OF が非常に悪い。以上のことより、この 3 試合が外れ値になったと考えられる。

6.2 分析結果 アウェー

表 2 重回帰分析結果 アウェー

	回帰係数	標準化残差	t-統計量	p-値
切片	-87.3575	2.6875	-32.505	2×10^{-16}
FGA	0.8390	0.0212	39.545	2×10^{-16}
FGM	0.1494	0.0027	59.387	2×10^{-16}
3PA	0.3550	0.0279	12.722	7.64×10^{-16}
3PM	0.0215	0.0013	15.738	2.48×10^{-16}
FTA	0.7224	0.0159	45.327	2×10^{-16}
FTM	0.0213	0.0015	14.096	5.01×10^{-15}
REB	0.0476	0.0219	2.166	0.0381

総当たり法を用いて変数選択を行ったところ、FGA,FGM,3PA,3PM,FTA,FTM,REB が説明変数となった。外れ値として出た 9,29 試合目を取り除き、重回帰分析を行った。得点に直接関係している変数以外で、REB が説明変数になっていることから、アウェーでは REB からのセカンドチャンスで得点に繋がっていると言える。外れ値である試合を見ると、9 試合目はシュート確率が良く、REB も多い事から、ほぼ完璧な試合で、29 試合目はシュート率が良すぎる。以上のことより、この 2 試

合が外れ値になったと考えられる。

6.3 分析結果 ホームとアウェー

表 3 重回帰分析結果 ホームとアウェー

	回帰係数	標準化残差	t-統計量	p-値
切片	-92.2327	2.7014	-34.142	2×10^{-16}
FGA	0.8926	0.0199	44.715	2×10^{-16}
FGM	0.1516	0.0024	60.884	2×10^{-16}
3PA	0.3637	0.0266	13.956	2×10^{-16}
3PM	0.0206	0.0012	16.120	2×10^{-16}
FTA	0.7475	0.0168	44.313	2×10^{-16}
FTM	0.0225	0.0014	15.367	2×10^{-16}
REB	0.0303	0.0400	0.758	0.451

得点の取り方にホームとアウェーで違いがあるかを分析するため、ホームに 1、アウェーに 0 を置いて、H/A のダミー変数を作り、11 変数で総当たり法を用いて変数選択を行った。その結果、H/A のダミー変数が選択されず、FGA,FGM,3PA,3PM,FTA,FTM,STL が説明変数となった。外れ値として出たホームの 3,4 試合目を取り除き、重回帰分析を行った。得点に直接関係している変数以外で、REB が説明変数になっていることから、REB からのセカンドチャンスで得点に繋がっていると言える。変数選択から、ホームとアウェーを合わせた場合の分析では、ホームとアウェーの違いは特に見られなかったが、ホームとアウェーを分けた場合の分析では、ホームでは DF から得点に繋がっていることが言え、アウェーでは OF でシュートを打った後の REB からのセカンドチャンスで得点に繋がっていることが言えるため、戦術面で違いがあると考えられる。外れ値である試合を見ると、3 試合目は OF が非常にいい、4 試合目は DF の良さが、OF に繋がっていない。以上のことより、この 2 試合が外れ値になったと考えられる。

7 おわりに

本研究を終えて、最高勝率チームではホームとアウェーで違いが出にくいことが分かった。しかし、その違いが出ないというところが、強いチームであり、負けないということでもあると思った。順位が真ん中のチーム、低いチームについても分析すれば、ホームとアウェーで違いが出ていたと思われるので、色々な強さのチームを含めて分析を行ったり、他の分析方法を用いて分析を行うことで、より明確にホームとアウェーの違いが現れるのではないかと思う。

参考文献

- [1] 青木繁伸：R による統計解析，オーム社，2009
- [2] 金明哲：R によるデータサイエンス，森北出版，2007
- [3] 中村永友：多次元データ解析法，共立出版，2009
- [4] ESPN-NBA：http://espn.go.com/nba/