

한국남자프로농구 경기기록 분석을 통한 승패결정요인 추정: 2010-2011시즌, 2011-2012시즌 정규리그 기록 적용

김세형¹ · 이준우² · 이미숙³

¹한국체육대학교 체육측정평가실 · ²호서대학교 기초과학연구소 · ³한국체육대학교 사회체육학과
접수 2012년 8월 31일, 수정 2012년 9월 18일, 게재확정 2012년 9월 23일

요약

한국남자프로농구 경기기록을 이용하여 승패결정요인을 분석하였다. 2010년 10월부터 2011년 3월까지, 2011년 10월부터 2012년 3월까지 치러진 정규리그 (540경기)의 기록을 분석하여 승패결정요인을 추정하였다. 한국농구연맹은 7개 공격변인과 7개 수비변인에 대한 자료를 제공하고 있다. 이들 자료 중에 공헌도와 공격력에 적용되는 6개 공격변인 (2점슛 성공률, 3점슛 성공률, 자유투 성공률, 공격리바운드, 어시스트, 턴오버)과 4개 수비변인 (수비리바운드, 스틸, 굿디펜스, 블록슛)이 승패에 미치는 영향을 통계적으로 분석하기 위해 로지스틱회귀분석과 의사결정나무분석을 적용하였다. 두 분석은 PASW와 Answer Tree 통계프로그램을 사용하였으며 모든 유의수준은 .05로 설정하였다. 로지스틱회귀분석 결과, 6개 공격변인 중 2점슛 성공률, 3점슛 성공률, 턴오버가 통계적으로 승패에 유의미한 영향을 미치고 4개 수비변인 중 굿디펜스를 제외한 수비리바운드, 스틸, 블록슛이 통계적으로 승패에 유의미한 영향을 미치는 것으로 나타났다. 그리고 공격변인 의사결정나무분석 결과에서는 2점슛 성공률이 51%-58%이며, 3P%가 31%를 초과하고 TO가 11개 이하일 때 승리할 수 있는 확률이 80.85%로 가장 높게 나타났다. 이에 반해 수비변인 의사결정나무분석 결과, 수비리바운드가 24개를 초과하고 스틸이 6개를 초과하며, 블록슛이 2개를 초과할 때 승리할 수 있는 확률이 94.12%로 가장 높게 나타났다.

주요용어: 로지스틱 회귀분석, 승패결정 요인, 의사결정나무분석, 한국농구연맹.

1. 머리말

프로농구에서 발전된 것은 매 경기마다 측정되는 경기기록이다. 프로농구 기록 관정원들에 관찰을 통해 측정되는 경기기록에 오차를 줄이는 교육이 이루어져 오면서 현재는 명확히 측정 기준을 통일하여 측정에 대한 타당성이 높아졌다 (Park, 2012). KBL은 인터넷 사이트를 통해 경기가 종료되고 24시간 이내에 출전선수별 14개의 변인 (7개의 공격변인과 7개의 수비변인)을 공개하고 있다 (박제영, 2005). 공격을 나타내는 7개 변인은 2점슛 성공률 (the rate of 2-Point shots; 2P%), 3점슛 성공률 (the rate of 3-Point shots; 3P%), 자유투 성공률 (the rate of Free Throw shots; FT%), 공격리바운드 (offensive rebound; OR), 어시스트 (assist; AS), 덩크슛 (dunk shots; DK), 그리고 턴오버 (turnover; TO)이다. 그리고 수비를 나타내는 7개 변인은 수비리바운드 (defensive rebound; DR), 스틸 (steal; ST), 굿디펜스 (good defence; GD), 블록슛 (block shots; BS), 자유투가 있는 파울수 (fouls with

¹ (138-749) 서울시 송파구 오륜동 88-15, 한국체육대학교 체육측정평가실, 박사.

² (336-795) 충남 아산시 배방읍 세출리 165, 호서대학교 기초과학연구소 (스포츠과학부 골프전공), 교수.

³ 교신저자: (138-749) 서울시 송파구 오륜동 88-15, 한국체육대학교 사회체육학과, 교수.

E-mail: misook@knsu.ac.kr

free throw; WFT), 자유투가 없는 파울수 (fouls without free throw; WOFT), 그리고 테크니컬파울 (technical fouls; TF)이다. 이 중에서 2P%, 3P%, FT%는 슈트도횟수와 성공횟수도 같이 공개한다.

예를 들어 2P% ($2P\% = 2P / 2PA \times 100$)는 2점슛 시도횟수 (2-Point shots Attempt; 2PA)와 2점슛 성공횟수 (success of 2-Point shots; 2P)에 의해 산출된다. 그동안 공개되었던 변인들을 적용시켜 적지않은 연구들이 승패결정요인에 대해 규명하였다 (Gu 등, 2009; Kim, 2008; Kim과 Park, 1999; Park 등, 2000; Park, 1997, 2001, 2003, 2004a, 2004b, 2005a, 2005b, 2007, 2008, 2009). 즉 승패집단 간의 통계적으로 유의한 차이가 나타나는 변인들을 분석한 정보를 제공하며, 14개의 변인들 중 어떤 변인이 승리에 통계적으로 어느 정도 유의한 영향을 미치는지를 분석하여 정보를 제공하였다. 그러나 수행된 선행연구들의 공통적인 특징은 변인에 적도를 정밀히 고려하지 않고 분석하였다.

구체적으로 프로농구 공격에 주요인이 되는 슈팅성공률 (2P%, 3P%, FT%)은 서열적으로 양적변수로 간주하는 데는 문제가 있다. 예를 들어 슈팅을 100번 시도하여 50번 성공한 선수의 슈팅성공률은 50%로 측정되지만, 슈팅을 2번 시도하여 1번 성공한 선수의 슈팅성공률에서도 50%로 측정된다. 이론적으로 슈팅성공률(%)은 슈트도횟수가 동일하지 않으면 사칙연산이 불가능한 질적변수이다 (Kim 등, 2011). 그럼에도 불구하고 한국프로농구 승패결정요인 분석에서 2P%, 3P%, FT%가 양적변수로 간주하여 분석해 왔다. 그 이유가 승패 집단간 슈트도횟수에는 큰 차이가 없다고 가정하였기 때문이다. 그러나 논리적으로 승패 집단간 슈트도횟수가 차이를 없다는 것에 대해 검증하지 않으면 연구의 타당화가 결여될 수 있다 (Kang, 2002; Kang, 2010; Wood 와 Zhu, 2006).

또한 한국프로농구 경기기록분석을 통한 승패결정요인 분석에 관련 연구들을 보면 공개되는 7개 공격변인 (2P%, 3P%, FT%, OR, AS, DK, TO)과 7개 수비변인 (DR, ST, GD, BS, WFT, WOFT, TF)을 모두 적용시켜 규명하였다. 그러나 공격변인에 DK (덩크슛), 수비변인에 WFT (자유투 유효파울수), WOFT (자유투 무 파울수), 그리고 TF (테크니컬파울)은 상대적으로 승리를 이끄는 공헌도와 경기력에 큰 영향을 미치지 못한다 (Park, 2012). 승리에 어느 정도 공헌을 하였는지, 어느 정도의 경기력을 갖추는지를 평가하는데 적용하고 있는 공식적인 평가모형에 DK, WFT, WOFT, TF 변인들은 적용되지 않고 있는 것이다. 특히 DK와 TF의 경우 한 경기에 이루어지지 않는 경우도 매우 많기 때문에 통계분석에 적용하기에 부적절한 변인이다.

따라서 이 연구는 DK를 제외한 6개의 공격변인 (2P%, 3P%, FT%, OR, AS, TO)과 WFT, WOFT, TF를 제외한 4개의 수비변인 (DR, ST, GD, BS)을 적용하여 승패결정요인을 통계적으로 분석하였다. 그동안 승패결정요인을 분석하는 다양한 추리통계법이 적용되어 왔다. 구체적으로 각 변인들이 승패에 미치는 영향을 분석하기 위해 판별분석 (discriminant)과 로지스틱회귀분석 (logistic regression), 그리고 데이터마이닝기법으로 의사결정나무분석 (decision tree analysis), 인공신경망분석 (artificial neural network analysis)이 주로 적용되어 왔다. 이 방법들은 통계적으로 각기 다른 분석조건을 지니고 있으며 결과를 산출하는 방법에 차이가 있다. 이 방법들 중에서 상대적으로 많이 적용되고 있는 로지스틱 회귀분석과 의사결정나무분석은 다음과 같은 장점을 지니고 있다.

로지스틱 회귀분석은 판별분석에 비해 분석이 이루어지기 위해 요구되는 기본가정이 엄격하지 않으며 다른 분석과 결합하여 다양한 정보를 줄 수 있는 확장의 가능성을 지니고 있다 (Hong, 2005). 또한 의사결정나무분석은 예측과 분류에서 다른 데이터마이닝기법에 비해 유용성이 높게 평가되고 있으며 다변량분석의 기본가정인 선형성, 정규분포성, 등분산성이 요구되지 않는다 (Kang와 Kim, 2009). 이러한 관점에서 이 연구에서는 로지스틱회귀분석과 의사결정나무분석을 적용하여 현재 한국프로농구 경기기록을 통한 승패결정요인을 추정하고자 한다. 승패에 영향을 미치는 경기력 결정요인을 통계적으로 파악하고 구조적으로 제시하여 선수와 감독들에게 승리를 위한 유의한 정보를 제공하는데 목적이 있다.

2. 연구방법

2.1. 연구자료

한국남자 프로농구 경기의 승리와 패배를 결정짓는 변인 (결정요인)을 분석하기 위해서 한국농구 연맹에서 공개하는 기록을 매 경기 종료 후 수집하였다. 2010년 10월 16일 부터 2011년 3월 20일까지 치러진 정규리그 270경기기록과 2011년 10월 15일 부터 2012년 3월 4일까지 치러진 정규리그 270 경기기록, 총 540경기기록을 연구의 자료로 사용하였다. Table 2.1은 연도별 팀 성적이다.

Table 2.1 Team result (2010-2011, 2011-2012 season)

2010-2011 season					2011-2012 season				
Rank	Team	Games	Victory	Defeat	Rank	Team	Games	Victory	Defeat
1	Busan	54	40	14	1	Wonju	54	44	10
2	Incheon	54	37	17	2	Anyang	54	35	19
3	Jeonju	54	34	20	3	Jeonju	54	31	23
4	Wonju	54	31	23	4	Busan	54	31	23
5	Seoul (Samsung)	54	28	26	5	Ulsan	54	29	25
6	Changwon	54	28	26	6	Incheon	54	26	28
7	Seoul (SK)	54	21	33	7	Changwon	54	21	33
8	Ulsan	54	20	34	8	Goyang	54	20	34
9	Anyang	54	17	37	9	Seoul (SK)	54	19	35
10	Goyang	54	15	39	10	Seoul (Samsung)	54	14	40

2.2. 연구변인

연구변인은 7개 공격변인 (2P%, 3P%, FT%, OR, AS, DK, TO)과 7개 수비변인 (DR, ST, GD, BS, WFT, WOFT, TF)으로 구성되었다. 구체적으로 공격변인에서 2P%는 제시되는 2점슛을 시도한 횟수 (2PA)와 2점슛을 성공한 횟수 (2P)에 의해 산출된 값이다. 3P%와 FT%도 동일하게 공개되고 산출된다. OR은 오픈스 리바운드로 경기 중 공격상황에서 실패된 슛팅에 리바운드를 성공한 빈도를 나타내며, AS는 어시스트로 슛팅성공을 유도한 패스에 빈도다. DK는 경기중 덩크슛을 성공시킨 빈도이며, TO는 턴오버로 공격팀이 필드골을 시도하지 못한 채로 볼의 소유권을 잃은 상황의 빈도가 측정된다. 그리고 7개 수비변인에서 DR은 디펜스 리바운드로 수비상황에서 상대팀이 실행한 슛팅이 실패했을 때에 성공한 리바운드 빈도를 나타낸다. ST는 스틸로 수비를 하다가 공격팀의 패스 또는 드리블 동작을 미리 예상하고 순간적으로 볼을 가로채는 것을 성공한 빈도가 기록된다. GD는 수비 선수가 공격자에게 TO를 범하게 해서 공격권을 가져오게 되었을 때 수비선수에게 GD 성공횟수가 기록된다. BS는 수비선수가 공격의 슛을 던진 상황에서 막는 빈도, WFT는 자유투가 있는 파울 빈도, WOFT는 자유투가 없는 파울 빈도를 나타낸다. TF는 비신사적인 행동으로 상대팀에게 자유투가 허용되는 상황에 기록된다.

이렇게 측정 및 계산되는 14개의 변인들 (공격변인 7개, 수비변인 7개) 중에 이 연구에서 프로농구 선수능력 평가에 적용되고 있는 공헌도와 경기력 모형 산출에 적용되는 변인들을 연구변인으로 적용하였다. 한국프로농구 공헌도와 경기력 모형은 Tabel 2.2 와 같다. 2PF는 2점슛 실패횟수, 3PF는 3점슛 실패횟수, 그리고 FTF는 자유투 실패횟수를 의미한다.

Table 2.2 Model of contribution point and match ability

Contribution point	$\{ \{ (\text{Total score} + \text{ST} + \text{BS} + \text{DR}) + (\text{OR} + \text{AS} + \text{GD}) \times 1.5 + \text{Time (min.)} / 4 \} - \{ (\text{TO} \times 1.5) + (2\text{PF} \times 1.0) + (3\text{PF} \times 0.9) + (\text{FTF} \times 0.8) \}$
Match ability	$\{ \{ (\text{Total score} + \text{ST} + \text{BS} + \text{DR}) + (\text{OR} + \text{AS} + \text{GD}) \times 1.5 \} / \text{Time (min.)} \times 40 - \{ (\text{TO} \times 1.5) + (2\text{PF} \times 1.0) + (3\text{PF} \times 0.9) + (\text{FTF} \times 0.8) \} / \text{Time (min.)} \times 40$

따라서 이 연구는 7개의 공격변인들 중에서 공헌도와 경기력 평가모형에 적용되지 않는 DK를 제외한 6개 공격변인 (2P%, 3P%, FT%, OR, AS, TO)과 7개 수비변인들 중에 공헌도와 경기력 평가모형에 적용되지 않은 WFT, WOFT, TF를 제외한 4개의 수비변인 (DR, ST, GD, BS)을 연구 변인으로 적용하였다.

2.3. 자료처리

한국프로농구 경기기록을 통해 승패결정요인을 통계적으로 추정하였다. 공격변인 중 슛팅성공률 (%)이 양적변인으로 간주할 수 있는지를 검증하기 위해서 승리한 집단과 패배한 집단간 슛팅시도 횟수가 통계적으로 유의한 차이가 없는지를 독립표본 t검증을 적용하여 검증하였다. 그리고 양적변인으로 검증된 슛팅성공률 (%)과 OR, AS, TO, 그리고 DR, ST, GD, BS 가 승패에 미치는 영향을 검증하기 위해 로지스틱 회귀분석을 실시하였다. 이 분석은 종속변수를 범주형으로 설계할 수 있는 회귀분석의 확장된 방법으로 동일한 설계방법을 가진 관별분석보다 광범위하게 적용되는 기법이다 (Hong, 2011). 로지스틱회귀분석은 PASW 프로그램을 사용하였으며 산출되는 exp(b) 값은 승산비 (오즈비: odd ratio)를 의미한다. exp 지수 (expotential)의 약자로 상수 2.718을 의미하며, b는 로지스틱 회귀계수다. 따라서 exp(b)는 2.718^b의 값이다. 이 연구에 로지스틱회귀분석 모형은 종속변인이 승·패 여부이고 독립변인은 공격변인 (2P%, 3P%, FT%, OR, AS, TO) 또는 수비변인 (DR, ST, GD, BS)이다.

의사결정나무 분석은 Answer Tree 프로그램을 사용하였다. 나무구조로 분류하는 통계방법으로 나무의 최상단에 뿌리마디 (root node), 나무 최하단에 잎마디 (leaf node), 그리고 뿌리마디와 잎마디 사이에 분류되는 중간마디 (internal node)로 구성된다. 중간마디는 분류가 이루어지는 부모마디 (parent node)와 분류되어지는 자식마디 (child node)로 구성된다. 이 연구에서 최소 사례수 (minimum number of case)는 부모마디에서 100, 자식마디에서 50으로 설정하였다. 선택한 알고리즘은 모형에 포함되어야 할 상호작용 효과를 고려하여 분석하는 CHAID (CHi-squared Automatic Interactive Detection) 알고리즘이며 χ^2 통계량의 p값을 분리기준으로 한다. 이 연구에서 실행한 모든 추리통계 분석의 유의수준은 .05로 설정하였다.

3. 결과

3.1. 독립표본 t검증

한국남자프로농구 경기기록 분석을 통해 승패에 영향을 미치는 변인 (요인)들을 추정하였다. Table 3.1은 이론적으로 질적변인인 슛팅성공률 (%)이 양적변인으로 간주할 수 있는지를 판단하기 위해 승패 집단간 슛팅시도횟수의 차이를 검증한 독립표본 t검증 결과이다. 승패집단간 2점슛 시도횟수 (2PA)는 t=1.305, p=.192 수준에서 통계적으로 유의한 차이가 없게 나타났다. 또한 3점슛 시도횟수 (3PA)는 t=-1.844, p=.073 수준에서 그리고 자유투 시도횟수 (FTA)는 t=1.999, p=.066 수준에서 모두 승패 집단간 통계적으로 유의한 차이가 없게 나타났다. 이러한 결과로 이 연구에서 한국남자프로농구 2점슛성공률 (2P%), 3점슛성공률 (3P%), 그리고 자유투성공률 (FT%)은 양적변인으로 간주할 수 있는 것을 알 수 있다.

Table 3.1 Difference of shooting frequency between victory and defeat

		N	M	SD	t	p
2PA	Victory	540	42.2	6.46	1.305	.192
	Defeat	540	41.5	6.19		
3PA	Victory	540	16.5	5.07	-1.844	.073
	Defeat	540	17.3	5.23		
FTA	Victory	540	16.2	6.12	1.999	.066
	Defeat	540	14.9	5.89		

3.2. 로지스틱회귀분석

한국프로농구 경기기록을 적용하여 승패결정요인을 추정하기 위해서 로지스틱 회귀분석을 실시하였다. Table 3.2 는 한국남자프로농구 공격변인 (2P%, 3P%, FT%, OR, AS, TO)이 승패에 미치는 영향을 분석하기 위한 단계선택방법에 따른 로지스틱회귀분석 결과이다.

Table 3.2 Estimating determinants of victory and defeat (offence variables)

	Step	b	SE	Wald	p	exp(b)
2P%	1	.113	.010	126.878	.001	1.119
3P%	2	.065	.006	99.458	.001	1.067
TO	3	-.152	.021	51.735	.001	.859

$$\text{logit}(\text{victory}) = -6.545 + .113 \times 2P\% + .065 \times 3P\% - .152 \times TO$$

b는 통계적으로 선택된 독립변수들의 로지스틱회귀계수이며, SE는 표본수를 고려한 표준오차를 나타낸다. Wald 값은 로지스틱회귀계수의 유의성을 검증하기 위해 로지스틱회귀계수 (b)를 표준오차 (SE)로 나누어 제공한 값으로 χ^2 분포 검증을 가능하게 한 통계치이다. p는 유의확률 값으로 이 연구에서 결정한 유의수준 (.05)과 통계적으로 유의성을 검증하기 위해 계산된 값이다. exp(b)는 승산비로 exp(b)=1일 때에 무효과, exp(b)>1일 때는 정적효과, 0<exp(b)<1일 때는 부적효과를 나타낸다 (Hong, 2011; Long, 1997).

6개의 공격변인 중 2P%, 3P%, 그리고 TO가 통계적으로 승패에 유의미한 영향을 미치는 것으로 나타났다. 1단계에서 2P%가 b=.113으로 Wald=126.878, p=.001 수준에서 통계적으로 유의하게 나타났으며 2단계에서는 3P%가 b=.065으로 Wald=99.458, p=.001 수준에서 그리고 3단계에서 TO가 b=-.152으로 Wald=51.735, p=.001 수준에서 통계적으로 유의하게 나타났다.

logit(승리)=-6.545+.113×2P%+.065×3P%-.152×TO는 로지스틱 회귀식으로 선택된 독립변수 (2P %, 3P%, TO)가 1단위 증가할 때에 승리할 수 있는 확률이 패배할 확률에 비해 높아지는 정도를 나타낸다. exp(b)값을 보면 2P%가 1.119, 3P%가 1.067, 그리고 TO가.859로 나타났다. 2P%, 3P%는 exp(b)값이 모두 1이상으로 2점슛, 3점슛 성공률이 높을수록 승리할 가능성이 높은 것을 알 수 있으며, TO는 exp(b)값이 1이하로 턴오버가 많을수록 승리할 가능성이 낮은 것을 알 수 있다.

Table 3.3 Classification accuracy (offence variable)

Observed	Expected			Accuracy (%)
	Victory	Defeat	Total	
Victory	384	152	536	71.6
Defeat	155	384	539	71.2
Total	539	536	1075	71.4

$$\chi^2 = 296.331, df = 3, p = .001$$

Table 3.3은 통계적으로 채택된 공격변인들 (2P%, 3P%, TO)을 독립변수로 하는 로지스틱회귀 분석 모형에 대한 통계적 유의성 및 분류의 정확도를 분석한 결과이다. 2점슛성공률, 3점슛성공률, 턴오버에 의해 승패를 예측하는 모형은 $\chi^2 = 296.331$, p=.001 수준에서 통계적으로 유의하게 나타났다. 승리와 패배에 대한 관찰치와 예측치 간의 차이를 보게 되면, 승리의 경우 전체의 71.6%가 패배의 경우에는 71.2%가 정확히 분류되어 전체적으로 71.4%의 정확도가 나타났다.

Table 3.4 Estimating determinants of victory and defeat (defence variables)

	Step	b	SE	Wald	p	exp(b)
DR	1	.190	.024	62.379	.001	1.210
ST	2	.225	.034	42.566	.001	1.252
BS	3	.139	.052	7.078	.001	1.149

$$\text{logit}(\text{victory}) = -5.384 + .190 \times DR + .225 \times ST + .139 \times BS$$

Table 3.4는 한국남자프로농구 수비변인 (DR, ST, GD, BS)이 승패에 미치는 영향을 분석하기 위한 단계선택방법에 따른 로지스틱회귀분석 결과이다. 4개의 수비변인 중 GD를 제외한 DR, ST, 그리고 BS가 통계적으로 승패에 유의미한 영향을 미치는 것으로 나타났다. 1단계에서 DR이 $b=.190$ 으로 $Wald=62.379$, $p=.001$ 수준에서 통계적으로 유의하게 승패에 영향을 미치는 것으로 나타났다. 2단계에서는 ST가 $b=.225$ 로 $Wald=42.566$ $p=.001$ 수준에서 유의하게, 3단계에서는 BS가 $b=.139$ 로 $Wald=7.078$, $p=.001$ 수준에서 통계적으로 유의하게 나타났다.

$logit(승리)=-5.384+.190 \times DR+.225 \times ST+.139 \times BS$ 는 로지스틱 회귀식으로 $exp(b)$ 값을 보면 DR이 1.210, ST가 1.252, 그리고 BS가 1.149로 나타났다. $exp(b)$ 값이 모두 1이상으로 수비리바운드, 스틸, 블록슛이 높을수록 승리할 가능성이 높은 것을 알 수 있다. 그리고 Table 3.5는 통계적으로 채택된 수비변인 들 (DR, ST, BS)을 독립변수로 하는 로지스틱회귀분석 모형에 대한 분류의 정확도를 분석한 결과다. DR, ST, BS에 의해 승패를 예측하는 모형은 $\chi^2 =117.059$, $p=.001$ 수준에서 통계적으로 유의하게 나타났다. 승리와 패배에 대한 관찰치와 예측치간의 차이를 보면 승리의 경우 전체의 65.5%가 패배의 경우는 72.1%가 정확히 분류되어 전체적으로 68.9%의 정확도가 나타났다.

Table 3.5 Classification accuracy (defence variable)

Observed	Expected			Accuracy (%)
	Victory	Defeat	Total	
Victory	199	105	304	65.5
Defeat	93	240	333	72.1
Total	292	345	637	68.9

$\chi^2 =117.059$, $df =3$, $p =.001$

3.3. 의사결정나무분석

한국남자프로농구 경기기록 분석을 통한 승패결정요인을 추정하기 위해 의사결정나무분석을 적용하였다. Figure 3.1은 2010-2011, 2011-2012시즌 프로농구 경기기록중 공격변인들 (2P%, 3P%, FT%, OR, AS, TO)이 승패에 미치는 영향력을 알아보기 위한 결과다.

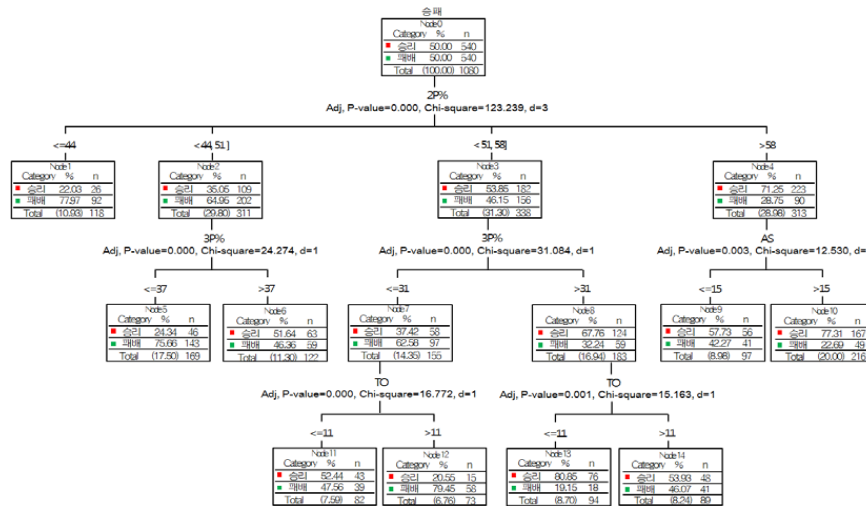


Figure 3.1 Result of decision tree analysis (offence variables)

공격변인 6개 중에 2P%, 3P%, TO, AS가 통계적으로 유의하게 나타났다. 2P% ($\chi^2 = 123.239$, $p = .001$)가 51%초과, 58%이하이며, 3P% ($\chi^2 = 31.084$, $p = .001$)가 31%를 초과하고 TO ($\chi^2 = 123.239$, $p = .001$)가 11개 이하일 때 승리할 확률이 80.85%로 가장 높게 나타났다. 반대로 2P%가 51%를 초과, 58%이하이며, 3P%가 31%이하이고 TO가 11개를 초과 할 때 승리할 확률이 20.55%로 가장 낮게 나타났다. Table 3.6은 의사결정나무모형 Figure 3.1에 분류의 타당성을 검증한 결과이다. 분류의 정확률은 69.71%, 오분류 확률은 30.29%로 나타났다.

Table 3.6 Classification accuracy of decision tree analysis (offence variable)

		Observed		Total
		Victory	Defeat	
Expected	Victory	410	208	618
	Defeat	130	332	462
	Total	540	540	1080

Classification accuracy (%)=69.71, Probability of misclassification (%)=30.29

Figure 3.2는 프로농구 수비변인들 (DR, ST, GD, BS)이 승패에 미치는 영향력을 알아보기 위한 의사결정나무분석 결과이다. 수비변인 4개 중에서 DR, ST, BS가 통계적으로 유의하게 나타났다. DR ($\chi^2 = 127.873$, $p = .001$)이 24개를 초과하고 ST ($\chi^2 = 10.893$, $p = .010$)이 6개를 초과하며, BS ($\chi^2 = 9.344$, $p = .038$)가 2개를 초과할 때 승리할 확률이 94.12%로 가장 높게 나타났다. 반대로 DR이 20개 이하이고, ST가 7개 이하일 때 승리할 확률이 29.44%로 가장 낮게 나타났다.

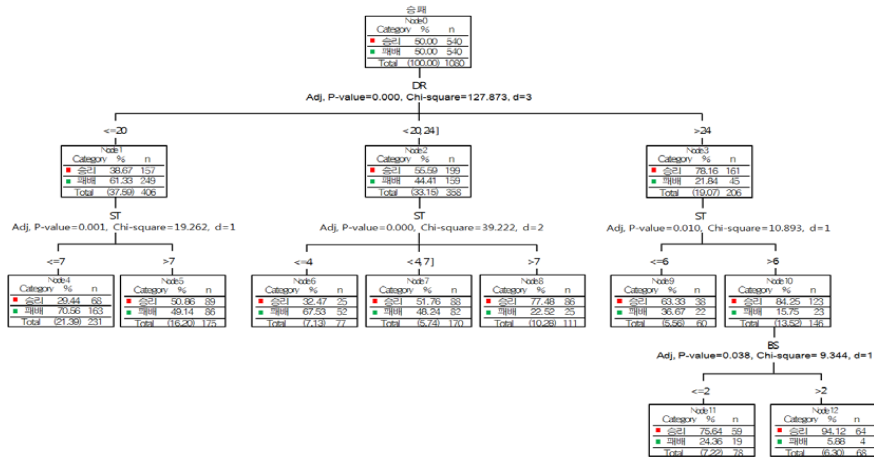


Figure 3.2 Result of decision tree analysis (defence variable)

Table 3.7은 의사결정나무모형 Figure 3.2에 분류의 타당성을 검증한 결과로, 분류의 정확률이 68.30%, 배반값인 오분류 확률은 31.70%로 나타났다.

Table 3.7 Classification accuracy of decision tree analysis (defence variable)

		Observed		Total
		Victory	Defeat	
Expected	Victory	424	238	662
	Defeat	116	302	418
	Total	540	540	1080

Classification accuracy (%)=68.30, Probability of misclassification (%)=31.70

4. 논의와 결론

최근 2년 동안 (2010-2011, 2011-2012시즌)의 한국남자프로농구 경기기록을 수집하여 승패결정요인을 분석하였다. 구체적으로 한국농구연맹에서 측정변인들을 확정하고 공개하는 기록들을 수집하여 통계적으로 승패에 영향을 미치는 요인들을 분석하였다. 기록되는 7개 공격변인들 (2P%, 3P%, FT%, OR, AS, DK, TO)과 7개 수비변인들 (DR, ST, GD, BS, WFT, WOFT, TF) 중에서 공식적인 공헌도와 경기력 평가모형에 적용되는 DK를 제외한 6개의 공격변인과 WFT, WOFT, TF를 제외한 4개의 수비변인을 적용하였다. 이 변인들이 승패에 미치는 영향을 통계적으로 분석하기 위해 로지스틱회귀분석과 의사결정나무 분석을 적용하였으며, 이에 대한 결과를 토대로 논의하면 다음과 같다.

그동안 적지 않은 스포츠 승패결정요인 분석 연구들은 승패에 주 결정요인이 되는 슛팅성공률 (%)을 양적변인으로 간주하였다. 다시 말하면, 승리한 팀과 패배한 팀은 모두 주어진 공격시간 내에 가능한 슛팅을 시도하기 때문에 슛시도횟수에 큰 차이가 없다고 가정한 것이다. 하지만 집단간 슛시도횟수에 유의한 차이가 있음에도 불구하고 슛성공률을 양적변수로 간주된다는 것은 연구 결과의 타당화가 결여되는 문제를 지니고 있다. 이러한 관점에서 이 연구에서는 승패 집단간의 슛팅시도횟수의 차이가 통계적으로 유의한지를 검증하였다. 즉 한국프로농구 경기기록 자료에서의 슛팅성공률 (2P%, 3P%, FT%)을 양적변인으로 적용하는데 큰 오차가 없다는 것을 통계적으로 검증하였다.

승패결정요인을 추정하기 위한 로지스틱회귀분석 결과를 보면 공격변인이 승패에 미치는 영향에서는 2점슛 성공률 (2P%), 3점슛 성공률 (3P%), 턴오버 (TO)가, 그리고 수비변인이 승패에 미치는 영향에서는 수비리바운드 (DR), 스틸 (ST), 블록슛 (BS)이 통계적으로 유의하게 나타났다. 그동안 한국프로농구 승패결정요인에 관해 많은 연구들을 실시한 Park (1997, 2001, 2003, 2004a, 2004b, 2005a, 2005b, 2007, 2008, 2009)의 연구들을 보면 7개의 공격변인들과 7개의 수비변인들을 각각 독립변인으로 설계하거나 동시에 14개의 변인을 독립변인으로 설계하여 승패에 미치는 영향을 판별분석을 적용하여 분석하였다. 본 연구와의 차이점이 있다면 경기당 거의 이루어지지 않는 TF와 DK, 그리고 실제 실수에 의한 파울과 전략적 파울로 행해지는 WFT와 WOFT를 독립변인으로 투입하였다는 것이 상이하다. 실제 본 연구 자료를 통해 빈도분석 실시해 보면, 테크니컬 파울은 540경기 중 456경기에서 전혀 이루어지지 않았으며, DK는 540경기 중 141경기에서 이루어지지 않았다. 다시 말해 통계적으로 다수의 결측치를 가지는 변인이다. 또한 자료처리방법에서 본 연구에서 적용한 로지스틱회귀분석이 아닌 판별분석을 적용했다는 것, 그리고 한 시즌 경기기록을 수집하여 분석하였다는 것과 두 시즌 경기기록을 수집하여 분석하였다는 차이점이 있다.

이처럼 연구설계에서 차이점은 있으나 본 연구결과와의 공통점은 공격변인에서 슛팅능력을 나타내는 2점슛 성공률 (2P%)과 3점슛 성공률 (3P%), 그리고 수비변인에서 수비리바운드 (DR)와 스틸 (ST)이 매년 공통적으로 승리에 유의한 변인으로 작용되고 있는 것이다. 그 외의 변인들은 본 연구결과에서 유의하게 나타난 턴오버 (TO), 블록슛 (BS) 이외에도 어시스트 (AS), 굿디펜스 (GD), 자유투성공률 (FT%)이 유의한 영향을 미치는 것을 보여주고 있다. 이러한 시즌별 승패결정요인 분석 결과의 차이점은 시즌마다 선수와 감독 교체와 연구 기록분석 결과로 중점이 되는 한 변인에 대한 집중 훈련으로 사료된다. 다시 말하면 매년 이루어지는 프로농구 경기기록분석 결과를 통해 어떤 변인이 승리에 유의하게 영향을 주는지에 대한 정보가 팀별 훈련에 영향을 준다고 할 수 있는 것이다.

또한 이 연구에서는 승패결정요인 분석 결과에 타당성을 높이기 위해 동일한 설계이지만 통계적으로 다른 추정 방법을 적용하는 의사결정나무분석을 사용하였다. 의사결정나무분석은 데이터마이닝 기법으로 많은 분석조건을 요구하고 있지 않으나 그만큼 많은 자료 (사례수)를 요구하는 분석으로,

최소 1000개 이상의 사례수로 분석 시 결과에 일반화가 가능하다(Lee, 2012; Choi, 2005). 즉, 사례수에 따라 최종결과가 다르게 나올 수 있는 것이다. 그동안 프로농구 한시즌 270경기기록 (540개)을 통해 의사결정나무분석을 적용한 연구들 (Gu 등, 2009; Kim, 2008)은 연구의 제한점으로 이 점을 제시하고 있다. 이러한 관점에서 본 연구에서는 최근 두 시즌 540 경기기록 (1080개)을 적용하였다.

그 결과 (Figure 3.1, 3.2)를 보면, 6개 공격변인 중 2점슛 성공률 (2P%), 3점슛 성공률 (3P%), 턴오버 (TO), 그리고 어시스트 (AS) 순으로 유의하게 채택되었다. 로지스틱회귀분석 결과와 동일하게 2P%, 3P%, TO의 중요성을 제시해 주었고 선수와 감독이 쉽게 이해할 수 있게 이 변인들이 어느 정도 일 때 승리할 확률을 그림화 시켜 제시해 주었다. 또한 로지스틱회귀분석에서 유의한 공격변인으로 나타나지 않은 AS에 중요성을 보여주고 있다. 두 분석에서 승리에 가장 높은 영향을 주는 2점슛 성공률이 어느 정도이고 어시스트가 어느 정도일 때 통계적으로 유의한 승리확률을 쉽게 보여주었다 (Figure 3.1).

4개의 수비변인 중에서는 DR, ST, BS 순으로 통계적으로 유의하게 나타났다. 이 연구에서 로지스틱회귀분석 결과와 동일하게 채택되었고, 모형을 그림화하여 (Figure 3.2) 보기 쉽게 보여주었다. 수비리바운드가 24개를 초과하고 스틸이 6개를 초과하며, 블록슛이 2개를 초과할 때 승리할 확률이 94.12%라는 것을 예측하였다. Kim과 Park (1999), Park 등 (2000)은 공격변인 (7개)과 수비변인 (7개)을 동시에 독립변인으로 투입하고 승패에 미치는 영향을 분석한 의사결정나무분석 결과에서는 DR이 승리에 가장 큰 영향을 주고 그 다음으로 2P%인 것을 공통되게 보여준다.

그동안 적지 않은 연구들이 한국프로농구 경기기록을 적용하여 승패결정요인을 추정해 왔다. 이 연구에서는 사전연구에서 고려하지 않은 공격능력을 나타내는 슈팅성공률 (%)의 척도를 검증하였으며 현재 프로농구 능력 (공헌도, 경기력)을 평가하는데 적용되고 있는 변인들을 통해 승패결정요인을 추정하였다. 즉 결과의 타당화를 높이기 위해서 추정방법이 다른 두 분석방법을 적용하였고 두 분석 결과에서 동일하게 프로농구 공격변인 중 2P%, 3P%, TO가 승리에 유의미한 영향력을 주는 것을 보여주었다. 이에 반해 수비변인 중에 DR, ST, BS이 승리를 위한 중요한 요인임을 확인하였다.

이 연구는 매년 한국프로농구협회에서 측정, 공개하는 경기기록 변인을 이용하여 통계적으로 승패에 유의한 영향을 주는 변인들을 제시하였다. 최근 2년간 이루어진 경기기록 자료를 적용하여 분석한 이 결과는 현재 한국남자프로농구의 횡단적인 정보가 될 수 있지만 외적타당화가 결여될 수 있다. 따라서 이 연구를 토대로 앞으로 누적되는 한국프로농구 경기기록 분석에 대한 다양한 연구가 지속적으로 수행된다면, 한국프로농구 팬들과 선수, 그리고 감독들에게 유용한 정보로 활용될 수 있을 것으로 기대된다.

참고문헌

- Choi, J. H. (2005). *Data mining*, Free Academy, Seoul.
- Gu, S. H. and Kim, H. S. and Jang, S. Y. (2009). A comparison study on the prediction models for the professional basketball games. *Journal of Korea Institute of Sport Science*, **20**, 704-711.
- Hong, S. H. (2005). *Bi-multi logistic regression*, Kyoyookbook, Seoul.
- Hong, S. H. (2011). *Bi-multi logistic regression program*, Korea University 2-season education measurement and evaluation workshop, Seoul National University, Seoul.
- Kang, B. S. and Kim K. S. (2009). *Statistics analysis of social science*, Hannarae, Seoul.
- Kang, S. J. (2002). *Sport research method and method of statistics*, winter season workshop, Korea Sport Measurement and Evaluation, Seoul.
- Kang, T. H. (2010). *Application of item response Theory*, winter season workshop, Korea Sport Measurement and Evaluation, Seoul.
- Kim, C. Y. and Park, J. Y. (1999). Analysis of contribution of won-lost factor in the 98-99 season Korean basketball league. *Journal of Korean Society of Sport and Leisure Studies*, **12**, 455-464.

- Kim, S. H. (2008). *Developing estimate model of victory and defeat through analyzing the record of the pro-basketball*, Master Thesis, Korea National Sport University, Seoul.
- Kim, S. H. and Kim, H. J. and Park, J. H. (2011). Development of model to evaluate handball shooting ability : Weight elicitation of shooting positions. *Korean Journal of Measurement and Evaluation in Physical Education and Sport Science*, **13**, 77-87.
- Lee, Y. G. (2012). *Data mining analysis*, Korea applied statistics workshop, Chung-Ang University, Seoul.
- Long, S. (1997). *Regression models for categorical and limited dependent variables*, Sage, Thousand Oaks, CA.
- Park, D. K. (2012). *Analyzing of contribution point in pro-basketball score*, winter season workshop, Korea Sport Measurement and Evaluation, Seoul.
- Park, J. Y. (1997). The analysis of the factor for winning a game in the 1997 season Korean basketball league. *Journal of Suwon University*, **15**, 311-318.
- Park, J. Y. (2001). The analysis of the factor for winning a game in the 2000-2001 season Korean basketball league. *Journal of Korean Society of Sport and Leisure Studies*, **16**, 1215-1224.
- Park, J. Y. (2003). The analysis of the factor for winning a game in the 2002-2003 season Korean basketball league. *Journal of Korean Alliance for Health Physical Education Recreation and Dance*, **42**, 793-893.
- Park, J. Y. (2004a). The analysis of the factor for winning a game in the 2003-2004 season Korean basketball league. *Journal of Korean Society for the Study Physical Education*, **9**, 185-195.
- Park, J. Y. (2004b). The analysis of the factors for winning a game in th 2004 winter season women's Korea basketball league. *Journal of Suwon University*, **22**, 601-611.
- Park, J. Y. (2005a). The analysis of the factor for winning a game in the 2004-2005 season Korean basketball league. *Journal of Suwon University*, **23**, 629-640.
- Park, J. Y. (2005b). The analysis of the factors for winning a game in the 28th Athens Olympic men's basketball game. *Journal of Korean Society of Sport and Leisure Studies*, **23**, 561-570.
- Park, J. Y. (2007). The analysis of the factors for winning a game in the 15th Asian men's basketball. *Journal of Korean Society of Sport and Leisure Studies*, **30**, 941-950.
- Park, J. Y. (2008). The analysis of the factor for winning a game in the 2000-2001 season Korean basketball league. *Journal of Korean Society of Sports Science*, **17**, 129-138.
- Park, J. Y. (2009). The analysis of the factors for winning a game in the 29th Beijing Olympic men's basketball game. *Journal of Korean Society of Sport and Leisure Studies*, **37**, 1425-1432.
- Park, J. Y. and Kim, C. Y. and Ji, E. B. (2000). Contribution of won-lost factor in th 99-2000 season Korean basketball league by decision tree analysis. *Korean Society of Sport and Leisure Studies*, **14**, 327-338.
- Wood T. M. & Zhu, W. (2006). *Measurement theory and practice in kinesiology*, Human Kinetics, Champaign, IL.

Estimating the determinants of victory and defeat through analyzing records of Korean pro-basketball

Sae Hyung Kim¹ · Jun Woo Lee² · Mi-Sook Lee³

¹Lab of Measurement and Evaluation in Physical Education, Korea National Sport University

²Basic Science Institute, Hoseo University

³School of Community Sport, Korea National Sport University

Received 31 August 2012, revised 18 September 2012, accepted 23 September 2012

Abstract

The purpose of this study was to estimate the determinants of victory and defeat through analyzing records of Korean men pro-basketball. Statistical models of victory and defeat were established by collecting present basketball records (2010-2011, 2011-2012 season). Korea Basketball League (KBL) informs records of every pro-basketball game data. The six offence variables (2P%, 3P%, FT%, OR, AS, TO), and the four defense variables (DR, ST, GD, BS) were used in this study. PASW program was used for logistic regression and Answer Tree program was used for the decision tree. All significance levels were set at .05. Major results were as follows. In the logistic regression, 2P%, 3P%, and TO were three offense variables significantly affecting victory and defeat, and DR, ST, and BS were three significant defense variables. Offensive variables 2P%, 3P%, TO, and AS are used in constructing the decision tree. The highest percentage of victory was 80.85% when 2P% was in 51%-58%, 3P% was more than 31 percent, and TO was less than 11 times. In the decision tree of the defence variables, the highest percentage of victory was 94.12% when DR was more than 24, ST was more than six, and BS was more than two times.

Keywords: Decision tree analysis, Korea Basketball League, logistic regression, 2010-2011, 2011-2012 season.

¹ Doctor, Lab of Measurement and Evaluation in Physical Education, Korea National Sport University, Seoul 138-749, Korea.

² Professor, Basic Science Institute (Department of Golf), Hoseo University, Chungnam 336-795, Korea.

³ Corresponding author: Professor, School of Community Sport, Korea National Sport University, Seoul 138-749, Korea. E-mail: misook@knsu.ac.kr