체육과학연구, 2014, 제25권, 제1호, 92-107.
Korean Journal of Sport Science
2014, Vol. 25, No. 1, 92-107.

한국프로야구 Post시즌 전력분석 및 예측모형

채진석 · 송종국*(경희대)

본 연구는 한국프로야구 포스트시즌 동안에 치러지는 준PO, PO, KS경기에서의 승, 패에 따른 서로 다른 형태의 경기력을 분석한 연구이다. 각 시리즈별 경기력의 차이점은 무엇인지를 제시하는 것이 1차 목적이며, 각 단기전별 최적의 예측모형을 제시하는 것이 2차 목적이다. 따라서 이러한 연구목적을 달성하기위해 1982년부터 2012년까지 실행된 준PO(준 플레이오프,1989-2012), PO(플레이오프, 1986-2012), KS(한국시리즈, 1982-2012)자료를 이용하여 각 단기전의 승리 팀 vs 패배 팀 간의 경기력변인에 따른 평균비교(t-test)를 하였으며, 예측변인으로 사용될 기술영역변인을 산출하기위해 동등화 과정을 거쳐 가중치(상관계수)를 사용 새로운 영역별(투수력, 타력, 수비력, 기동력, 득점 및 집중력, 선구능력)변인을 생성하여 이 들 변인을 이용 3가지 예측모형(판별모형, 이항로지스틱회귀모형, 인공신경망모형)에 투입하여 예측모형의 정확도를 제시하였다. 준PO와 KS에서는 실책이 유의한 변인으로 나타났으며 PO에서는 삼루타가 준PO와 KS에서는 이루타가 유의미한 변인으로 새롭게 나타났다. 또한 승/패에 큰 영향을 주는 기술영역변인은 준PO에서는 투수력, PO에서는 타력, KS에서는 투수력으로 나타났다. 준PO에서의 최종예측모형은 예측변인이 제일 적게 투입되면서 적중률이 가장 높은 인공신경망모형(통계적 기준: 4개)이 최종선택 되었고, PO에서는 전문가 기준에 의해 만든 기술영역변인을 예측변인으로 적용한 인공신경망 모형(전문가 기준:5개)이 다른 두 모형보다는 예측 적중률이 더 좋았다. KS에서의 특징은 모든 변수가 적용되어 예측변인을 만든 인공신경망모형(모든 변수: 6개)이 최종모형으로 선정 되었다. 또한 준PO, PO, KS의 종합적인 적중률을 살펴보면 판별모형보다는 이항로지스틱회귀모형이, 이항로지스틱회귀모형 보다는 인공신경망모형이 더 우세한 분류정확성을 보이고 있었다.

주요어: 한국프로야구, Post시즌, 판별분석, 로지스틱 회귀분석, 인공신경망분석

서 론

1982년 창단 첫해, 6개 팀으로 시작한 한국프로야구는 86년에는 7개 팀으로, 91년에는 8개 팀으로, 2013년에는 9개 팀으로 증가 하였고, 2015년에는 10개 팀으로 운용할 것으로 확정된 가운데 선수들의 경기력은 많은 발전이 있어왔다. 특히 분석도구의 향상과 기록 자료를 활용하려는 마인드의 변화로 경험(經驗)과 감(感)의 야구에서 통계를 활용한 과학적인 야구로 변모 하였다. 초창기 한국프로야구의 전력분석은 기록지를 분석하여 기술통계를 제시하는 수준이였으나 최근에는 모든 팀이 컴퓨터 프로그램을 이용 실시간 상대 팀 정보를 제시하는 수준까지 와 있다. 특히 한국시리즈에서 2연패를 한 삼성이나 6년 연속 한국시리즈에 진출하여 3번 우승 3번 준우승을 한 SK를 보아도 전력분석을 중요시하는 팀들은 좋은 성적을 나타내고 있다. 이들 팀들은 시합 전에 전력분석 미팅을 갖으며, 상대팀의 장단점이 기록된 의미 있는 통계분석자료를 제공받고 있다. 또한 경기 중에도 전력분석원의 분석 자료는 실시간 제공 되고 있다.

이와 같이 경기장에서 생산된 자료를 이용해 의미 있는 경기력지수로 변환하여 선수들에게 제공하는 팀의 전력분석원은 현재 9개 구단 모두 전력분석팀이라는 이름으로 존재하고 있다. 전력분석의 중요한 또 하나의 예로

논문 투고일 : 2013. 10 .07.
논문 수정일 : 2014. 02. 06.
게재 확정일 : 2014. 02. 19.
* 저자 연락처 : 송종국(jksong@khu.ac.kr).
* 이 논문은 2011년 한국연구재단의 지원을 받아 수행된 연구임 (354-2011-1-G00082).

2013년 WBC(World Baseball Classic)경기에 대비한 한국대표팀과 NC 다이노스와의 연습경기에서 대만의 전력분석요원이 심판후보생으로 잠입해 한국의 전력을 탐색한 것이 밝혀져 공식 사과를 하였다는 언론보도가 있었다(이대호, 2013). 이와 같이 전력분석은 현대의 스포츠경기에서 승리를 위한 필수 요인으로 인식되고 있다(김주학, 2007; 김세형 등 , 2008; 최형준, 2009).

또한 이러한 인식을 기초로 야구현장에서 발생하는 모든 자료를 경기력 측면에서 분석함에 있어 새롭게 분석하려는 연구자들도 나타났으며, 그들을 미국에서는 세이버메트리션 이라고 부른다. 이 용어는 1971년 설립된 미국 야구연구 협회의 (Society for American Baseball Research) 머리글자를 따서 SABR라 칭하였고, 컴퓨터 시뮬레이션과 복잡한 고등수학의 도움을 빌려 야구기록을 분석하고 전통적인 야구이론을 발전시키는 일을 하는 야구통계의 신조어로서 경기력 분석측면에서 본다면 새로운 야구 통계분석 방법을 일컫는 용어이다(Costa et al, 2008; 채진석, 2011). 이러한 미국의 세이버들의 연구회와 비슷한 연구학회가 한국에서도 2013년 6월 1일 한국야구연구학회(Society for Korean Baseball Research)란 이름으로 창립되었다.

이렇게 세이버메트리션이 유명해진 계기는 미국 프로야구 메이저리그(MLB) 30개 구단 중 가장 가난한 구단 '오클랜드 어슬레틱스'가 4년 연속 포스트시즌에 진출이라는 성공 신화를 담은 책이 발표된 이후이다(Michael, 2003). MLB의 오클랜드 단장은 세이버메트리스적 통계에 기반을 둔 선수평가 기법을 도입하여 타율보다는 출루율과 장타율에 초점을 맞추고 팀 전체를 혁신한 결과 2000년부터 4년 연속 Post시즌에 진출 했고 지난 5년 동안 승수가 양키즈구단을 제외한 가장 많은 구단으로 기록되었다. 이러한 결과는 경기력을 과학적인 통계방법으로 분석하려는 세이버들의 연구가 활발해졌기 때문이다(Thorn & Palmer, 1984; Total Baseball, 2003; Levernier & Barilla, 2006; James, 2008).

이와 같은 세이버메트리스적 통계적 방법은 MLB뿐 아니라 KLB(한국프로야구)에서도 경기력분석과 아울러 통계적 예측모형들에 관련된 연구들이 매년 발표되고 있었다(최용석과 심회정, 1995; 장인식과 원정심, 1996; 홍석미 등, 1996; 박진, 1999; 최옥진, 2002; 김승대, 2003; 조영석과 조용주, 2003, 2004, 2005a, ; 이장택과 김용태, 2005, 2006a, 2006b; 황서영, 2007; 신상근 등, 2007; 박승현, 2008; 홍종선과 최정민, 2008; 채진석과 엄한주, 2010; 채진석, 2011; 최영근과 김형문, 2011; 승희배과 강기훈, 2012; 천영진과 최형준, 2013).

지금까지 프로야구자료를 이용하여 예측변인을 활용한 승, 패 예측 논문들은 미래에 발생될 수 있는 결과를 예측하는데 적용되는 여러 가지 통계적 방법을 사용 하였다. 그 중에서 프로야구분석에서 가장 빈번하게 사용되고 있는 예측기법은 선형회귀분석(Linear regression analysis), 곡선추정(Curve Estimation), 판별함수분석(discriminant function analysis), 로지스틱 회귀분석(logistic regression analysis, Daniel, 2005), 주성분 회귀분석(Principal component regression analysis), 분류나무분석(Classification tree analysis)과 최근에 빈번하게 사용되는 인공신경망 분석(artificial neural network analysis)등을 들 수 있다(서재순과 정태충, 1993; 허준희와 정태충, 1998; 정태충, 1999; 김차용, 2001; 오광모와 이장택, 2003; 조영석과 조용주, 2005b; 이영훈, 2007; 채진석 등, 2010; 배재영 등, 2012).

이상과 같이 야구에 관련한 선행연구가 다양한 관점에서 다뤄지고 있었고 또한 앞선 연구 중에서도 정규리그 자료만을 이용한 연구(채진석, 등 2010)가 있었으나 본 연구와의 큰 차이점은 종속변수가 서로 다르며, 자료의 특성이 서로 다르다. 즉, 정규리그의 종속변수는 Post시즌을 진출 한 팀과 진출 못한 팀으로 범주화 하였고 Post시즌에서는 승리한 팀과 패한 팀으로 이분화 하였다. 또한 장기전인 정규시즌 동안은 구단별 최근 133개 경기를 치르는 것에 비해 단기전은 3전 2선승제나 5전 3선승제, 7전 4선승제로 진행되기 때문에 투수운용과 타자, 수비수 기용 등 팀 운용 면에서 서로 달아 승/패에 미치는 변인들의 특성이 차이가 있을 것이다. 또한 Post시즌 기간 내의 각 단기전인 준 플레이오프, 플레이오프, 한국시리즈에서도 서로 다른 특징적인 경기가 나타날 것으로 예상된다.

따라서 본 연구에서는 단기전이라고 할 수 있는 Post시즌(준PO, PO, KS)자료만을 이용하여 팀의 승, 패에 미치는 주요한 경기력변인이 무엇이고, 어떻게 서로 다른지 알아보는 것이며, 어떤 팀이 승리 할 것인지 3가지 예측모형을 만들어 가장 예측 정확성이 좋은 모형을 선택하고, 선택된 모형을 이용 다음 시합 때 어느 팀이 승

표 1. 분석변인

단기전		기술영역변인	독립변인(예측변인)	종속변인
P O S T 시즌	준 PO, PO, KS	투수와 관련된 변인	방어율(평균자책점,ERA), 세이브, 투구이닝, 피타자, 피안타, 피홈런, 피사4구, 탈삼진, 실점, WHIP, 자책점, 탈삼진당피사4구.	1. 승 0. 패
		타격과 관련된 변인	타율, 타수, 득점, 일루타, 이루타, 삼루타, 홈런, 루타, 타점, 타점율, 타득점, 도루, 사4구, 삼진, 출루율, 장타율, 순장타율, OPS, 삼진당사4구, 실책, 도루자 도루성공율, PSN.	
		관련변인	연도, 경기수, 팀명, 승수, 패수, 승률.	

리 할 것인지 가능성을 제시 할 수 있다. 이러한 일련의 연구목적은 전력의 차이가 미세한 팀 간의 Post시즌 승부에서 매 단기전별로 어떤 측정변인을 더 비중 있게 연습해야하며, 어떤 기술영역을 더 보강해야 하는지를 방법론적 측면에서 현장에 제시 할 수 있다는 것과, 예측모형을 이용 팀의 승, 패를 가늠 할 수 있어 이 또한 사전대비에 효과가 있을 것으로 예상한다.

연구방법

연구대상

본 연구대상은 프로야구 정규리그가 끝난 이후의 Post시즌기간에 치러지는 준PO와 PO 그리고 KS의 단기전자료이며 프로야구 정규시즌이 끝나면상위 4팀은 Post시즌에 진출하게 된다. Post시즌은 준PO(정규시즌 3,4위전), PO(준 플레이오프 승자와 정규시즌2위 간 전), KS(플레이오프 승자와 정규시즌 1위 팀 간 대결)로 구성된다.

연구대상의 측정 자료는 준PO(준플레이오프, 1989-2012년), PO(플레이오프, 1986-2012년), KS(한국시리즈, 1982-2012년)경기의 진출 팀과 패배 팀의 경기기록변인을 의미한다(표 1).

본 연구의 분석에 사용된 분석변인인 〈표 2〉는 준PO, PO, KS에서 기록된 변수를 기초로 MLB(미국프로야구)에서 사용하거나 SABR (Society for American Baseball Research)에서 제시한 공식에 의해 생성된

표 2. 기술영역별 합성변인 생성과정

변인	합성변인 생성과정	비고
방어율	(자책점×9)/투구이닝	ERA(평균자책점)
WHIP	(피안타+피사4구)/투구이닝	이닝당출루허용율
피사 4구	피死(사)구+ 피4구	몸에맞는볼+볼넷허용
탈삼진당피사 4구	피사4구/탈삼진	수치가 적을수록 좋음
출루율	(안타+4사구)/(타수+4사구+희비)	OBP
타율	안타/타수	AVG
삼진당사4구	사4구/삼진	BB/K
OPS	장타율 + 출루율	수치가 클수록 좋음
루타	단타+ 이루타×2 + 삼루타×3+홈런×4	TB
장타율	루타/타수	SLG
순장 타율	장타율 - 타율	ISOP
호타 준족	(2×홈런×도루)/(홈런+도루)	PSN
타점율	타점/타수	RBI%
타득점	타점+득점	수치가 클수록 좋음
실책	실책수	수치가 적을수록 좋음
도루 성공율	도루/(도루+도실)	SBP
사4구	4구 + 사(死)구	볼을 잘 보는 능력

합성변수 또는 비율변수로 행위에 대한 성공/실패 또는 효율성 지수를 제공한다 (채진석과 엄한주, 2010).

자료수집 및 처리방법

본 연구의 분석에 사용된 경기기록 변인들은 스포츠투 아이(주)에서 제공한 원 자료를 바탕으로 Microsoft Office Excel 2007프로그램을 이용하여 다양한 합성변인 및 비율변인들을 생성하였고, 일부 한국프로야구 기록대백과(2009년)와 한국프로야구 연감(2009-2012)을 이용하였다. 또한 현장 전문가(감독, 코치)의 의견을 묻는 설문지(채진석, 2011)도 이용 되었다. 자료처리 분석 방법에 있어서는 본 논문의 의미 있는 결과를 도출하기위해 자료의 탐색을 거쳐 빈도분석, 독립 t검정, 상관분석, 판별분석, 이항로지스틱회귀분석, 인공신경망 분석을 SPSS Windows 18.0프로그램에 적용하여 결과를 산출하였다.

예측모형의 예측변인으로 사용한 기술영역변인(투수력, 타력, 기동력, 수비력, 선구능력, 득점 및 집중력) 생성과정은 다음과 같다. Post시즌 동안 측정한 측정변수 및 합성변수, 비율변수(표 2)들을 3가지(전체변수, 전문가 기준:80%, 통계적 기준:p<.05)선정기준에 따라 선택된 측정변수들은 단위가 서로 다른 변수들이므로 척도동등화과정을 거쳐 각각의 측정변수와 승률과의 상관계수를 가중치로 하여 〈수식 1〉과 같이 산출하였다 (채진석, 2011).

자료처리순서는 다음과 같다. 각각의 경기력변수(35개)가 기술영역변인으로 변환 되는 기준은 3가지 측면을 고려하였다. 3가지 선정기준 중 첫 번째는 전문가(51명)의 의견이 80% 이상(아주 큰 영향을 준다+큰 영향을 준다)인 변수만을 선정한 결과 5개의 기술영역변인을 산출 하였다(표 3). 두 번째는 승리집단과 패배집단간의 평균비교를 통해 유의미한(p<.05) 변인만을 적용하여 4개의 기술영역변인을 산출 하였다(표 2). 마지막으로 세 번째는 연구변인 모두를 적용하여 6개의 기술영역변인들로 변환하였다. 이렇게 3가지 측면에 의해 변환된 기술영역변인들을 예측변수(독립변수, 설명변수)로 하여 3가지 예측모형각각 투입하면 준PO, PO, KS전에서의 승·패 가능성을 제시할 수 있었다.

또한 3가지 분석모형들의 수리적 알고리즘은 다음과 같다. 판별모형은 종속변수(반응변수)에는 집단구분을, 독립변수(설명변수, 예측변수)에는 집단을 설명하는 변수를 사용하게 된다. 이모형을 이용하기 위해서는 각 개체는 여러 개의 집단 중에서 어느 집단에 속해있는지 알려져 있어야 하며, 과거 이미 알려진 집단의 각각의 경우에 대하여 측정된 변수 들을 이용하여 각 집단들을 가장 잘 구분할 수 있는 판별함수식($Z_i = \alpha_0 + \beta_{i1}x_1 + \beta_{i2}x_2 + \cdots + \beta_{ip}x_p$)을 찾아내어 이 식을 이용 각 집단(종속변수)에 속한 개인의 판별점수를 산출하여, 각 개체들이 어느 범주에 속하는지 분류하고 예측할 목적으로 사용되는 통계기법이다(Anderson, 1958; Cooley, 1971; 古俗野 恒과 김명수, 2003; 양병화, 2006).

또한 로지스틱회귀분석은 독립변수에는 명목, 서열, 등간, 비율척도로 측정된변수를 사용할 수 있으나 종속변수에는 오직 명목척도로 측정된 범주형 변수만을 사용하여 개별 관측 치들이 어느 집단에 속하는지를 예측한다. 이 로지스틱기법의 수학적함수식은

$$E(Y|X) = P(X) = \frac{e^{b_0 + b_1X_1 + b_2X_2 + \cdots + b_nX_n}}{1 + e^{b_0 + b_1X_1 + b_2X_2 + \cdots + b_nX_n}}$$

가 되며, X를 이용하여 예측한 Y값은 $E(Y|X)$이고 Y가 이산변수일 때 $E(Y|X)$는 확률의 개념을 가지므로 $P(X)$로 나타낼 수 있다. 이모형은 선형함수가 아닌 S자 곡선인 로지스틱 함수로 상한계가 1이고 하한계가 0인 관계로 선형함수로 표현할 수 없어 분석하는데 문제가 있

수식 1.

$$Y_{l,k} = \sum_{i \in C_{k^+}} \frac{x_{l,i} - m_i}{M_i - m_i} \times r_i + \sum_{j \in C_{k^-}} \frac{x_{l,i} - M_j}{M_j - m_j} \times r_j$$

$$T_l = \sum_{k=1}^{n} Y_{l,k} \left(\begin{array}{l} C_{k^+} = k\text{번째 기술영역에서 상관계수가 + 인 지수들의 집합}, \\ C_{k^-} = k\text{번째기술영역에서 상관계수가 - 인 지수들의 집합}. \\ l = l\text{번째 구단}, n = \text{기술영역의 갯수}, \left| C_{k^+} \cup C_{k^-} \right| = k\text{번째기술영역의 크기}. \end{array} \right)$$

표 3. 전문가가 선정한 승·패에 미치는 영향력 (단위: %, n=51)

기술영역	경기력변인	아주 큰 영향을 준다	큰 영향을 준다	작은 영향을 준다	아주 작은 영향을 준다	80%이상 순위
투수력	방어율	74.5	25.5	0	0	100
	세이브	62.7	37.3	0	0	100
	WHIP	54.9	43.1	2	0	98
	자책점	39.2	51	9.8	0	90.2
	피홈런	39.2	47.1	9.8	3.9	86.3
	피사 4구	27.5	47.1	21.6	3.9	74.6
	피안타	11.8	60.8	23.5	3.9	72.6
	투구이닝	23.5	47.1	17.6	11.8	70.6
	탈삼진	15.7	51	29.4	3.9	66.7
	피타자	11.8	33.3	39.2	15.7	45.1
타력	홈런	66.7	31.4	2	0	98.1
	OPS	25.5	70.6	3.9	0	96.1
	출루율	52.9	41.2	5.9	0	94.1
	삼루타	41.2	51	5.9	2	92.2
	이루타	29.4	60.8	7.8	2	90.2
	PSN	29.4	58.8	11.8	0	88.2
	장타율	27.5	54.9	15.7	2	82.4
	타율	37.3	39.2	21.6	2	76.5
	일루타	21.6	54.9	21.6	2	76.5
	루타	21.6	54.9	15.7	7.8	76.5
	순장타율	11.8	64.7	23.5	0	76.5
	타수	17.6	29.4	37.3	15.7	47
득점·집중력	타점	64.7	31.4	3.9	0	96.1
	타점율	64.7	31.4	3.9	0	96.1
	득점	58.8	35.3	5.9	0	94.1
수비력	실점	51	43.1	2	3.9	94.1
	실책	52.9	39.2	3.9	3.9	92.1
기동력	도루	17.6	66.7	15.7	0	84.3
	도루성공율	25.5	58.8	13.7	2	84.3
	도루자	15.7	45.1	29.4	9.8	60.8
선구능력	4구	31.4	43.1	23.5	2	74.5
	사 4구	13.7	49	33.3	3.9	62.7
	삼진	7.8	51	29.4	11.8	58.8
	사구	13.7	35.3	39.2	11.8	49

다. 그러므로 이 확률을 로짓(logit)으로 변환하면 확률의 상, 하 한계가 없어지며, 독립변수와 로짓의 관계를 선형함수[$\ln\left(\frac{p}{1-p}\right)=\alpha_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\cdots+\beta_nx_n$]로 표현이 가능해져 일반적인 선형(linear) 회귀분석을 사용할 수 있게 되는 것이다.

따라서 이 로짓선형함수식의 좌변인 자연로그진수 값인 괄호 안은 승산비로 분자인 p는 개별 개체들이 어느 집단에 속할 확률이고 분모인 1-p는 그 집단에 속하지 않을 확률이므로 우변의 n개의 예측변수(X)를 이용하여 계산한 결과, logit값이 클수록 그 집단에 속할 확률이 높다는 것이다. (Agresti, 2002; Hosmer, & Lemeshow. 2000; 홍세희, 2005). 끝으로 인공신경망분석(artificial neural network analysis)의 원리는 인간이 학습(learning)을 통하여, 다음의 행동을 행하는 거와 같이 이를 컴퓨터에 학습용 자료를 이용하여, 가장 최적의 결과를 학습시키고, 새로운 자료 또는 상황에 그 학습의 결과를 응용하여, 예상 결과를 도출하게 하는 것이다. 본 연구의 신경망구조는 입력층(input layer), 은닉층(hindden layer), 출력층(output layer) 등 3개 층으로 구성되어 있고 각 층은 몇 개씩의 뉴런(neuron, 신경세포)을 포함하고 있다. 입력층의 뉴런이 자극(각종 자료)을 받으면 은닉층의 신경세포가 이를 전달받아 선형 결합($L = w_1X_1 + w_2X_2 + \cdots + w_nX_n$, w는 가중치)으로 연결되고, 이 값이 커질수록 뉴런(신경세포)이 활성화되고, 반대의 경우 비활성화가 된다. 이 활성화 값의 정도를 S라고 하면 S가 제한된 범위 ($0 \le S \le 1$, $-1 \le S \le 1$)를 취하도록, L로 부터 S로의 변환 $S = f(L)$에 활성함수(로지스틱함수: $S = \frac{e^L}{1+e^L}$, 쌍곡 탄젠트함수: $S = \frac{e^L - e^{-L}}{e^L + e^{-L}}$)가 개입된다. 출력노드는 은닉 뉴런으로부터 오는 신호들을 가중치로 결합하여 최종 반응을 내는데 목표변수가 연속형 일 때는 신호들의 가중치 결합을 그대로 적용하나, 범주형인 경우는 각 범주별 출력 값이 모두 확률 값으로 제시하는 소프트맥스 변환이 적용된다. 여기서 k는 $O_k = \exp(L_k) / \sum_{n=1}^{K} \exp(L_n), k = 1,2,\cdots K$)출력범주를 나타내는 인덱스이고 K는 출력 범주 수이다(이용구와 허준, 1999; 허명회와 이용구, 2003; 허명회, 2008).

연구결과

Post시즌은 준 플레이오프(정규시즌 3,4위전), 플레이오프(준 플레이오프 승자와 정규시즌2위 간 전), 한국시리즈(플레이오프 승자와 정규시즌 1위)로 정의하며, 통계분석을 통해 각 단기전 별 유의한 경기력변인을 산출하고 분석하여 승리를 위해서는 어떤 변인이 중요한지 제시하였으며, 또한 경기력변수인 측정변수를 상대평가가 가능한 기술영역별 상대지수(투수력, 타력, 수비력, 득점 및 집중력, 기동력, 선구능력)로 변환하여 어떤 지수가 승, 패에 더 크게 작용하는지 알아보았다. 또한 이들 기술영역별 상대지수를 예측변인으로 예측모형을 제시하였다.

1. 준PO 진출 팀의 경기력분석

준PO에서 경기력변수들에 대한 승/패 팀 간의 평균비교를 한 결과인 〈표 4〉를 살펴보면 유의미한 차이가 큰 변수의 순위는 세이브(t=5.51), 타점율(t=3.56), OPS(t=3.370), 장타율(t=3.368),방어율(t=-3.32), 순장타율(t=2.954)순으로 나타났으며 모든 변수를 이용해서 산출한 기술영역상대지수에서는 투수력(t=5.092)이 승, 패에 가장 큰 차이를 보였고, 그 다음은 타력(t=3.227), 득점·집중력(t=2.868), 수비력(t=2.619) 순으로 나타났다. 그러나 기동력과 선구능력은 평균의 차이가 없는 것으로 나타났다.

또한 각 기술영역지수들과 승률간의 관련성 정도를 알아본 결과 〈표 5〉에 나타난 바와 같이 3가지 기준 모두 투수력이 승률과 가장 강한 양의 상관관계를 나타내고 있으며, 그 다음이 타력, 득점 및 집중력, 수비력 순으로 정적인 관련성을 보이고 있다. 특히 종합전력과 승률과의 관계에서는 3가지 기준에서 모두 정적인 관련성이 아주 강한 것으로 나타났지만, 전문가 기준(r=.832**)에서 다소 높은 것으로 나타났다.

2. PO 진출 팀의 경기력분석

1986년부터 2012년까지의 PO에 진출한 팀을 승리팀과 패전 팀으로 구분하여 각 기록변수에 대하여 평균을 비교한 결과는 〈표 6〉에 나타내었다. 유의미한 차이가 큰 변수들은 OPS(t=5.621), 장타율(t=5.508), 순장타율(t=4.739),출루율(t=4.689), WHIP(t=-4.689), 방어율(t=-4.596), 타율(t=4.170), 타점율(t=3.708) 순으로 나타났으며 기술영역상대지수에서는 타력(t=5.117), 투수력(t=4.920), 득점 및 집중력(t=3.447), 수비력(t=3.080), 선구능력(t=2.294)에서 유의미한 평균의 차이가 있었고 기동력에서만 차이가 없었다. 또한 각 기술영역지수들과 승률간의 관련성 정도를 알아본

표 4. 준PO 승리 팀 vs 패전 팀의 평균 비교 및 상관분석 (1989-2012년, n=44)

기술영역	경기력변수	승리 팀		패전 팀		t	p	상관계수 (승률)
		M	SD	M	SD			
투수력	방어율	2.84	1.34	4.85	2.49	-3.32	.002	-.558**
	WHIP	1.24	.284	1.53	.406	-2.71	.01	-.482**
	세이브	1.10	.700	.19	.402	5.51	.000	.606**
	투구이닝	27.4	9.32	26.6	9.37	.269	.789	.042
	피타자	112.5	40.9	117.7	42.5	-.415	.680	-.069
	피안타	24.6	12.5	27.5	11.5	-.816	.419	-.166
	자책점	9.09	6.25	13.6	6.85	-2.30	.027	-.385*
	피홈런	1.86	1.61	2.86	1.89	-1.89	.066	-.324*
	피사 4구	10.2	5.60	13.0	7.20	-1.43	.161	-.213
	탈삼진	18.68	7.0	18.73	7.8	-.020	.984	.032
	탈삼진당피사4구	.578	.311	.713	.339	-1.38	.175	-.249
타력	타율	.273	.051	.238	.047	2.396	.021	.474**
	타수	101	36.8	99.7	35.9	.120	.905	.018
	일루타	19.8	9.6	19.0	9.5	.284	.778	.088
	이루타	4.55	2.81	2.86	2.50	2.12	.040	.292
	삼루타	.27	.55	.45	.74	-.926	.487	-.156
	홈런	2.86	1.89	1.86	1.61	1.891	.066	.324*
	루타	41.2	15.6	33.6	18.1	1.50	.141	.261
	출루율	.353	.063	.306	.056	2.622	.012	.474**
	장타율	.415	.091	.325	.085	3.368	.002	.565**
	순장타율	.141	.069	.087	.051	2.954	.005	.469**
	OPS	.768	.137	.631	.132	3.370	.002	.571**
	PSN	1.98	1.51	1.35	1.38	1.447	.155	.212
득점·집중력	득점	15.1	7.71	9.91	7.02	2.311	.026	.385*
	타점	13.6	6.99	9.05	6.32	2.286	.027	.377*
	타득점	28.7	14.6	19.0	13.3	2.31	.026	.382*
	타점율	.137	.059	.084	.037	3.56	.001	.584**
수비력	실책	1.55	1.71	2.68	1.89	-2.09	.042	-.290*
	실점	9.91	7.02	15.1	7.71	-2.31	.026	-.384*
기동력	도루	2.05	1.76	1.82	1.44	.470	.641	.015
	도루성공율	.556	.361	.497	.321	.574	.569	-.030
	도루자	1.43	1.50	1.48	1.40	-.109	.914	.077
선구능력	사 4구	13.0	7.21	10.2	5.59	1.43	.161	.213
	삼진	18.7	7.8	18.4	6.56	.146	.884	-.020
	삼진당사4구	.713	.339	.579	.310	1.37	.178	.249
기술영역 상대지수	투수력	2.129	.378	1.482	.461	5.092	.0001	.667**
	타력	1.794	.562	1.243	.610	3.227	.002	.539**
	득점·집중력	.843	.372	.542	.320	2.868	.006	.473**
	수비력	.479	.141	.364	.150	2.619	.012	.398**
	기동력	.035	.016	.034	.016	.226	.822	.041
	선구능력	.190	.105	.145	.089	1.542	.131	.256
	팀 종합전력	5.505	.824	3.827	.830	6.725	.0001	.820**

*P < .05, **P < .01, ***p < .001

표 5. 준PO전의 변수선택 기준에 따른 상관분석(기술영역변인 vs 승률)

항목	승률		
	전체 변수	통계적 기준	전문가 기준
투수력	.667**	.707**	.707**
타력	.539**	.566**	.547**
득점·집중력	.473**	.473**	.496**
수비력	.398**	.398**	.398**
기동력	.041		-.028
선구능력	.256		
종합전력	.820**	.826**	.832**

*P<.05, **P<.01

표 6. PO 승리 팀 vs 패전 팀 간의 평균 비교 및 상관분석 (1986-2012년, n=58)

기술영역	경기력변인	승리 팀		패전 팀		t	p	상관계수 (승률)
		M	SD	M	SD			
투수력	방어율	2.73	1.45	4.41	1.34	-4.596	.001	-.707**
	WHIP	1.24	.25	1.53	.23	-4.689	.001	-.624**
	세이브	1.38	1.08	.66	.77	2.936	.005	.469**
	투구이닝	40.42	10.3	39.40	10.8	.368	.714	.057
	피타자	167.7	48.7	172.5	47.5	-.382	.704	-.052
	피안타	34.8	12.6	39.9	12.2	-1.567	.123	-.204
	자책점	13.3	9.16	18.8	7.02	-2.591	.012	-.417**
	피홈런	2.62	3.23	4.21	2.14	-2.201	.032	-.344*
	피사4구	16.3	7.67	19.97	7.23	-1.867	.067	-.252
	탈삼진	28.1	11.2	26.1	9.7	.703	.485	.115
	탈삼진당피사4구	.64	.45	.82	.26	-1.79	.079	-.371**
타력	타율	.272	.036	.232	.036	4.170	.001	.564**
	타수	147.3	40.3	147.0	40.6	.026	.979	-.002
	일루타	27.4	8.83	26.3	8.6	.497	.621	.058
	이루타	7.21	3.87	5.55	3.75	1.655	.103	.224
	삼루타	1.07	1.03	.34	.55	3.329	.002	.234
	홈런	4.21	2.14	2.62	3.23	2.201	.032	.343**
	루타	61.9	18.8	48.9	22.9	2.352	.022	.310*
	출루율	.358	.038	.305	.046	4.689	.001	.626**
	장타율	.423	.067	.319	.077	5.508	.001	.685**
	순장타율	.151	.049	.087	.055	4.739	.001	.617**
	OPS	.781	.098	.624	.114	5.621	.001	.699**
	PSN	3.12	1.74	2.13	2.24	1.876	.066	.284*
득점·집중력	득점	21.3	7.43	14.5	9.26	3.111	.003	.442**
	타점	19.6	6.57	13.8	9.12	2.776	.007	.427**
	타득점	40.9	13.9	28.2	18.4	2.959	.005	.436**
	타점율	.133	.034	.091	.048	3.708	.001	.618**
수비력	실책	3.21	2.83	2.93	1.83	.440	.661	-.003
	실점	14.41	9.3	21.21	7.36	-3.09	.003	-.443**
기동력	도루	3.83	2.99	2.97	2.68	1.157	.252	.181
	도루성공율	.621	.307	.584	.344	.433	.666	.110
	도루자	1.82	1.40	1.29	1.32	1.013	.316	.125
선구능력	사4구	19.83	7.23	16.2	7.66	1.852	.069	.252
	삼진	26.4	9.90	28.0	11.1	-.588	.559	-.111
	삼진당사4구	.808	.265	.641	.453	1.709	.093	.365**
기술영역 상대지수	투수력	2.060	.50	1.480	.393	4.920	.0001	.698**
	타력	2.657	.624	1.637	.874	5.117	.0001	.644**
	득점·집중력	1.094	.308	.733	.472	3.447	.001	.530**
	수비력	.293	2.83	.216	1.83	3.080	.003	.442**
	기동력	.206	.084	.197	.079	.384	.702	.093
	선구능력	.286	.072	.228	.116	2.294	.026	.389**
	팀종합전력	6.619	.919	4.535	1.411	6.625	.0001	.825**

*P<.05, **P<.01, ***p<.001

표 7. PO전의 변수선택 기준에 따른 상관분석(기술영역변인 vs 승률)

항목	승률		
	전체 변수	통계적 기준	전문가 기준
투수력	.698**	.725**	.725**
타력	.644**	.661**	.599**
득점·집중력	.530**	.530**	.552**
수비력	.442**	.442**	.442**
기동력	.093		.167
선구능력	.389**		
종합전력	.825**	.852**	859**

*P<.05, **P<.01

결과 〈표 7〉에 나타난바와 같이 3가지 기준 모두 투수력이 승률과 가장 강한 양의 상관관계를 나타내고 있으며, 그 다음이 타력, 득점 및 집중력, 수비력 순으로 강한 양의 관련성을 보이고 있다. 특히 종합전력과 승률과의 관계에서는 3가지 기준에서 모두 정적인 관련성이 아주 강한 것으로 나타났지만, 전문가 기준(r=.859**)에서 다소 높은 것으로 나타났다(표 7).

3. KS 진출 팀의 경기력분석

한국시리즈에서의 승리 팀과 패전 팀 간의 경기력변수들의 평균비교는 〈표 8〉에 잘 나타냈으며 유의미한 차이가 큰 변인들은 세이브(t=5.79), 타점율(t=4.86),

표 8. KS 승리 팀 vs 패전 팀 간의 평균 비교 및 상관분석 (1982-2012년, n=60)

기술영역	경기력변인	승리 팀		패전 팀		t	p	상관계수 (승률)
		M	SD	M	SD			
투수력	방어율	2.65	1.31	4.11	1.42	-4.15	.0001	-.556**
	WHIP	1.23	.214	1.44	.241	-3.63	.0001	-.453**
	세이브	1.90	1.03	.57	.728	5.79	.0001	.637**
	투구이닝	51.4	11.0	50.2	11.3	.438	.663	.058
	피타자	211.4	48.2	216.5	44.8	-.427	.671	-.057
	피안타	41.6	12.4	46.9	10.7	-1.77	.081	-.264*
	자책점	15.6	8.91	22.3	7.57	-3.14	.003	-.413**
	피홈런	3.17	2.14	3.80	2.12	-1.15	.254	-.244
	피사 4구	22.1	8.20	24.6	7.93	-1.22	.229	-.087
	탈삼진	38.6	12.8	33.3	13.3	1.57	.122	.156
	탈삼진당피사4구	.619	.278	.808	.289	-2.58	.012	-.282*
타력	타율	.255	.039	.223	.031	3.48	.001	.507**
	타수	185.4	39.1	184.9	41.7	.045	.965	.019
	일루타	34.0	8.42	31.5	9.07	1.11	.273	.186
	이루타	8.10	3.48	6.00	2.64	2.64	.011	.311*
	삼루타	.93	.94	.87	1.14	.25	.806	.080
	홈런	3.80	2.12	3.17	2.14	1.15	.254	.212
	루타	68.2	16.6	58.8	20.0	1.99	.052	.303*
	출루율	.342	.040	.305	.037	3.65	.0001	.493**
	장타율	.371	.064	.312	.054	3.85	.0001	.555**
	순장타율	.117	.039	.089	.033	2.93	.005	.444**
	OPS	.713	.098	.618	.083	4.06	.0001	.562**
	PSN	2.06	.570	1.55	.722	3.07	.003	.368
득점·집중력	득점	25.33	7.60	17.63	9.86	3.39	.001	.443**
	타점	23.43	7.09	16.33	8.94	3.41	.001	.464**
	타득점	48.77	14.6	33.97	18.8	3.41	.001	.454**
	타점율	.129	.037	.084	.034	4.86	.0001	.641**
수비력	실책	3.60	2.13	5.27	2.07	-3.08	.003	-.374**
	실점	17.63	9.86	25.33	7.60	-3.39	.001	-.459**
기동력	도루	5.17	3.88	4.20	2.27	1.18	.243	.167
	도루성공율	.593	.240	.606	.214	-.21	.831	.043
	도루자	3.14	1.59	2.70	1.62	1.08	.284	.123
선구능력	사4구	24.57	7.91	22.3	8.38	1.08	.286	.127
	삼진	33.33	13.3	38.63	12.8	-1.57	.122	-.192
	삼진당사4구	.807	.29	.624	.278	2.50	.015	.335**
기술영역 상대지수	투수력	1.905	.389	1.346	.319	6.085	.0001	.661**
	타력	2.059	.639	1.432	.677	3.689	.0001	.526**
	득점·집중력	.921	.326	.539	.400	4.055	.0001	.538**
	수비력	.560	.168	.416	.122	3.820	.0001	.478**
	기동력	.141	.063	.144	.047	-.214	.831	.004
	선구능력	.350	.106	.279	.106	2.592	.012	.331**
	팀종합전력	5.936	.825	4.156	.967	7.673	.0001	.803**

*P<.05, **P<.01, ***p<.001

표 9. KS전의 변수선택 기준에 따른 상관분석(기술영역변인 vs 승률)

항목	승률		
	전체 변수	통계적 기준	전문가 기준
투수력	.661**	.666**	.688**
타력	.526**	.546**	.549**
득점·집중력	.538**	.538**	.559**
수비력	.478**	.478**	.478**
기동력	.004		.150
선구능력	.331**	.335**	
종합전력	.803**	.817**	.861**

*P < .05, ** P <.01

표 10. 예측모형 검정

단기전		준PO		PO		KS	
모형	적중률 기준	모형검정	적중률	모형검정	적중률	모형검정	적중률
판별 모형	통계적 기준(4)	χ^2:42.0 df:4, p<.001	93.2%	χ^2:46.7 df:4, p<.001	87.9%	χ^2:59.3 df:5, p<.001	86.7%
	전문가 기준(5)	χ^2:41.38 df:5, p<.001	90.9%	χ^2:44.7 df:5, p<.001	89.7%	χ^2:60.2 df:5, p<.001	88.3%
	전체 변수(6)	χ^2:38.71 df:6, p<.001	90.9%	χ^2:46.0 df:6, p<.001	89.7%	χ^2:59.9 df:6, p<.001	91.7%
이항 로지스틱 회귀분석 모형	통계적 기준(4)	χ^2:53.9 df:4, p<.001	97.7%	χ^2:55.0 df:4, p<.001	91.4%	χ^2:58.6 df:5, p<.001	90.0%
	전문가 기준(5)	χ^2:61.0 df:5, p<.001	100%	χ^2:51.1 df:5, p<.001	87.9%	χ^2:58.4 df:5, p<.001	90.0%
	전체 변수(6)	χ^2:51.8 df:6 p<.001	95.5%	χ^2:55.2 df:6 p<.001	87.9%	χ^2:59.6 df:6 p<.001	93.3%
인공 신경망 모형	통계적 기준(4)	ROC곡선아래 면적 승리 1: 1.00 패배 2: 1.00	100%	ROC곡선 아래 면적 승리1:.976 패배2:.976	93.1%	ROC곡선아래 면적 승리 1:1.00 패배 2 1.00	93.3%
	전문가 기준(5)	ROC곡선아래 면적 승리 1: 1.00 패배 2: 1.00	100%	ROC곡선 아래 면적 승리1:1.00 패배2:1.00	98.3%	ROC곡선아래 면적 승리 1:1.00 패배 2 1.00	98.3%
	전체 변수(6)	ROC곡선아래 면적 승리 1: 1.00 패배 2: 1.00	100%	ROC곡선 아래 면적 승리1:1.00 패배2:1.00	96.5%	ROC곡선아래 면적 승리 1:1.00 패배 2 1.00	100%

방어율(t=-4.15), OPS(t=4.06),장타율(t=3.85), 출루율(t=3.65), WHIP(t=-3.63)으로 나타났다. 기술영역상대지수에서는 투수력(6.085)이 승, 패 간에 가장 큰 평균의 차이를 보였으며, 그 다음은 득점 및 집중력(t=4.055), 수비력(t=3.820), 타력(t=3.689), 선구능력(t=2.592)순으로 유의미한 차이를 보였고, 기동력에서만 차이가 없었다. 또한 각 기술영역지수들과 승률간의 관련성 정도를 알아 본 결과 〈표 9〉에 제시한 바와 같이 KS에서는 선구능력이 유의미한 결과를 보였고, 3가지 기준 모두 투수력이 승률과 가장 강한 양의 상관관계(r=.661**, r=.666**, r=.668**)를 나타내고 있었다. 그 다음이 득점·집중력, 타력, 수비력, 선구

능력 순으로 강한 관련성을 보인 것은 전체변수를 사용한 경우와 전문가 기준에 따른 경우다. 특히 종합전력과 승률과의 관계에서는 3가지 기준에서 모두 정적인 관련성이 아주 강한 것으로 나타났지만, 전문가 기준(r=.861**)에서 다소 높은 것으로 나타났다.

4. Post Season의 예측모형 선정

Post시즌 내 3개의 단기전 중 첫 번째인 준PO전의 예측모형비교에서는 9가지모형 모두 양호한 것으로 판명 되었으며(p< .01, AUC=ROC 곡선아래면적>.70), 9가지모형에 의한 각 사례의 분류적중률을 간명성에 입각하여 살펴보면 모든 변수를 예측변수(6개)로 적용한 적중률이 오히려 예측변인이 적게 투입된 통계적 기준(4개)과 전문가 기준(5개)보다 다소 떨어지는 결과가 나타났다. 그러므로 예측변인이 제일 적게 투입되면서 적중률이 가장 높은 인공신경망모형(통계적 기준: 4개)이 최종선택 되었다(표 11). 대체적으로 적중률을 살펴보면 판별모형 보다는 로지스틱회귀모형이, 이항로지스틱모형 보다는 인공신경망모형이 더 우세한 분류정확성을 보이고 있었다. 또한 최종 선정된 예측모형을 이용 2012년에 치려진 실제경기인 준PO전의 롯데와 두산 경기를 비교 분석한 결과 〈표 11〉에서 나타난 바와 같이 롯데가 승리 할 것으로 예측한 것이 옳은 것으로 나타났다. 두 번째 단기전인 PO전의 예측모형비교에서는 9가지모형 모두 모형검증에서는 양호한 것으로 판명 되었으며(p<.01, AUC=ROC 곡선아래면적>.70), 9가지모형에 의한 각 사례의 분류적중률을 살펴보면 비슷하나 인공신경망모형들이 다소 높았고 최종모형으로는 전문가 기준(예측변인:5개)의 인공신경망모형(98.3%)을 선택하였다. 또한 최종 선정된 예측모형을 이용 2012년에 치려진 실제경기인 SK와 롯데 경기를 비교 분석한 결과 〈표 11〉에서 나타난 바와 같이 SK가 승리 할 것으로 예측한 것이 옳은 것으로 나타났다.

또한 KS전의 예측모형비교에서는 9가지 모형 모두 모형검증에서는 양호한 것으로 판명 되었으며(p<.01, AUC=ROC 곡선아래면적>.70), 9가지모형에 의한 각 사례의 분류적중률을 살펴보면 비슷하나 판별모형<이항로지스틱회귀모형<인공신경망모형 순으로 다소 적중률이 높은 것으로 나타났다(표 10). 또한 최종모형으로는 예측변인이 6개가 투입된 인공신경망모형(100%)으로 정하였으며, 이렇게 정한 최종모형을 이용 2012년에 치려진 실제경기인 삼성과 SK 경기를 비교 분석한 결과 삼성이 승리 할 것으로 예측하였고 그 결과도 같은 것으로 나타났다(표 11).

5. 종합분석

전체 측정변수 모두를 사용하여 만든 기술영역별 변인 별로 승·패간 평균차이 검정한 결과의 t값으로 방사형 그래프인 〈그림 1〉를 나타내었고, t의 절댓값이 크면 클수록 승/패 간에 차이가 큰 것을 알 수 있다. 그 결과로 준PO에서의 승/패에 따른 각 기술영역별 크기 순위는 투수력> 타력> 득점 및 집중력> 수비력 순으로 나타났다. PO에서는 타력> 투수력> 득점 및 집중력> 수비력> 선구능력 순으로, KS에서는 투수력> 득점 및 집중력> 수비력> 타력> 선구능력 순으로 나타났다. 특히 PO에서의 특징은 타력이 가장 승/패 간에 큰 차이를 보였고, KS에서는 타력 보다는 득점·집중력과 수비력에서 승/패간 차이가 더 컸다. 이는 각 시리즈별로 승리 팀과 패전 팀 간의 경기력 변인에서 차이가 있음을 알 수 있었다(그림 2). 또한 3가지 변수선택 기준에 따른 최종 예

표 11. 최종 예측모형 선택 및 실제 자료 적용

단기전	준PO					PO					KS				
변수선택	통계적 기준					전문가 기준					모든 측정변수				
예측변수	투수력, 타력, 득점·집중력, 수비력					투수력, 타력, 득점·집중력, 수비력, 기동력					투수력, 타력, 득점·집중력, 수비력, 기동력, 선구능력				
최종 Model	인공신경망모형					인공신경망모형					인공신경망모형				
실제적용	팀	승/패	실제결과	예측확률	예측결과	팀	승/패	실제결과	예측확률	예측결과	팀	승/패	실제결과	예측확률	예측결과
2012년	롯데	3/1	승	.999	승	SK	3/2	승	.858	승	삼성	4/2	승	.955	승
	두산	1/3	패	.001	패	롯데	2/3	패	.142	패	SK	2/4	패	.045	패

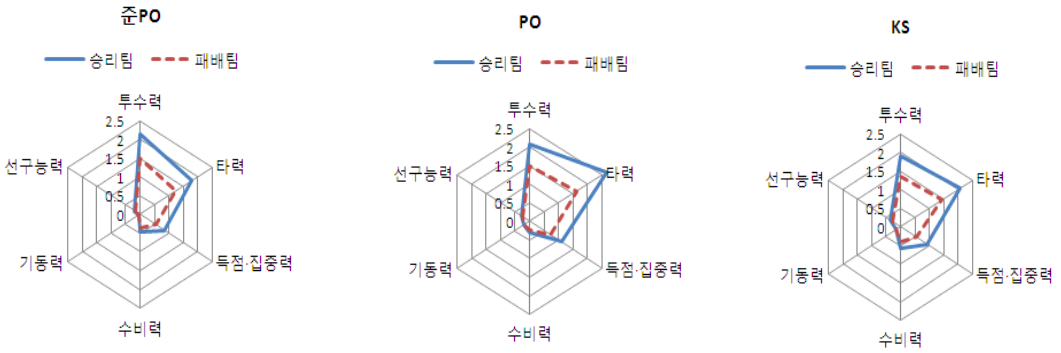

그림 2. Post시즌 전력비교

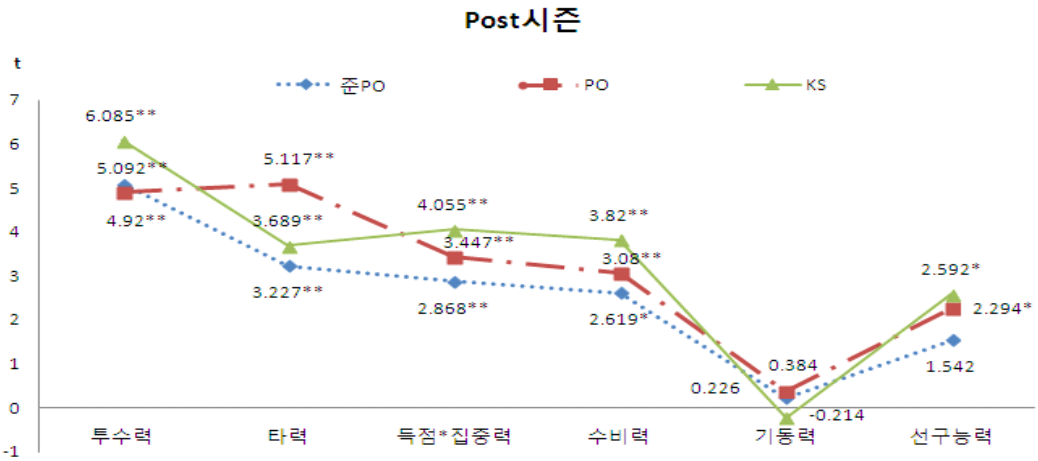

그림 3. 각 단기전에서 기술영역변인의 중요도

측모형을 살펴보면, 영역별 상대평가지수을 예측변수로 하여 만든 3가지 예측모형을 간명의 원칙에 따라 제시한 결과, 준PO에서는 통계적 기준(승/패 간의 평균비교에서 유의미한 변수로 선택하여 만든 4개 예측변인이 투입된 인공신경모형이 선정 되었고, PO에서는 전문가 의견에 따른 예측변인으로 적용한 인공신경망 모형이 다른 두 모형보다는 예측 적중률이 더 좋았다. 또한 KS에서는 모든 변수가 적용된 인공신경망이 최종모형으로 선정되었다.

논 의

본 연구는 포스트시즌자료만을 이용, 경기력 분석 및 예측모형을 제시한 논문으로 PS시즌 동안 치려지는 준PO, PO, KS의 차이점이 무엇이며 승리를 위해서는 어떤 경기력변수가 더 중요한지 제시 하였다. 또한 각 구단의 과거성적을 이용 어느 팀이 단기전에 승리 할 것인지의 가능성을 제시 할 수 있는 예측모형을 산출하기위해 최초 측정변수를 3가지(통계적 기준, 전문가기준, 전체 연구변수)에 따라 선정하였고, 선정된 변수에 가중치(상관계수)를 적용, 기술영역변인(투수력, 타력, 수비력, 득점 및 집중력, 선구능력)을 생성한 후 이들 변인을 예측변인으로 하여 예측모형을 산출하였다.

과거 정규리그 자료(1998년)만을 이용, 로지스틱회귀분석과 의사결정나무분석을 사용하여 유의미한 측정변수만을 모형에 참여 시키는 방법으로 승·패 예측모형을 제시한 김차용(2001)은 이항로지스틱회귀분석을 통해 승, 패에 가장 중요하게 미치는 요인은 집중율(타점/안타수), 총루타, 실책수 등으로 나타났으며, 예측모형의 분류 적중률은 이항로지스틱모형이 78.3%, 의사결정나무분석이 78%를 옳게 분류 하였다고 하였다. 또한

정규시즌 자료만을 이용한 채진석 등(2010)의 연구에서는 투수력과 타력, 득점력 변인에서 포스트 진출 팀과 진출하지 못한 팀 간의 뚜렷한 차이를 보였다고 하였으며, 예측모형 비교에서는 판별분석과 로지스틱분석이 일부 분석에서 각각의 기법이 선택한 예측변인은 다른 경우가 있었지만 대부분의 분석에서는 동일한 변인을 예측변인으로 채택하였고 예측적중률 또한 거의 동일하였으나 동일한 변인을 예측변인으로 선택하여 분석한 인공신경망분석의 예측적중률은 판별과 로지스틱분석보다는 다소 높았으나 큰 차이는 없었다고 하였다. 또한 이영훈(2007)은 정규리그 자료를 이용하여 분석한 결과 출루율이 장타율에 비해 3배 정도 중요하며 작전야구가 팀의 득점 및 실점에 주는 영향은 부정적이라고 하였다.

그러나 본 연구는 정규시즌이 아닌 포스트시즌자료를 연구대상으로 하였으므로 선행연구들과 대상의 차이와 종속변인의 차이가 있지만 투·타의 조화와 수비력의 중요성은 일치함을 나타내고 있으며, 인공신경망모형이 판별모형과 로지스틱회귀모형보다는 다소 분류 적중률이 정확한 면이 나타나고 있었다는 것은 일치하고 있었다(최형준과 김주학, 2006).

본 연구의 특징은 승/패에 있어 준PO와 KS에서는 실책이 유의미한 변인으로 나타났으며, PO에서는 삼루타가 준PO와 KS에서는 이루타가 유의미한 변인으로 새롭게 나타났다. 또한 승/패에 큰 영향을 주는 기술영역변인은 준PO에서는 투수력, PO에서는 타력, KS에서는 투수력으로 나타났다.

결 론

본 연구결과를 종합한 결론을 크게 2가지 측면에서 제시하려한다. 첫 번째는 경기력 측면으로서 준PO에서의 승/패 간의 기술영역변인의 유의미한 차이는 투수력〉 타력〉 득점 · 집중력〉 수비력 순으로 나타났고 투수력과 관련된 측정변수인 세이브, 방어율, 이닝 당 출루 허용율(WHIP)에서 승/패 간에 큰 경기력 차이를 보였다. 세이브 차이의 의미는 승리 팀은 패전 팀 보다 마무리 투수의 역할이 강했고 방어율과 WHIP차이는 선발, 중간투수의 역할도 강했다는 의미다. 또한 타력과 득점 및 집중력에 관련된 변수들 중에서 OPS(장타율+출루율)와 타점율이 가장 큰 차이를 보인 것은 상대 팀보다 득점이 많아야 이기는 경기이므로 장타로 출루를 하였으면 출루한 선수를 홈으로 불러들이는 것이 득점의 가장 중요 요소이므로 이런 점에서 승리 팀은 패전 팀 보다 타력과 집중력에서 더욱 앞서 있었다. 더욱이 실책에서 승/패 간의 유의미한 차이를 보인 것은 수비력의 중요성을 나타내고 있으며, 수비력과 투수력 간의 강한 상관관계(r=.640**)가 존재하고 준PO에서 팀 승리를 위해서는 수비를 바탕으로 한 투수력과 타력의 조화에 집중력을 강화한 플레이를 제시한다.

PO에서의 승/패 간의 경기력의 특징은 준PO와 KS와는 달이, 타력〉 투수력〉 득점 및 집중력〉 수비력〉 선구능력 순으로 유의미한 차이를 나타냈다. 투수력보다 타력(OPS, 순장타율, 타율)에서 경기력 차이를 보인 것이 큰 특징이다. PO에서의 승리 팀은 패전 팀 보다 삼루타와 홈런에서 유의미한 차이를 보인 반면 준PO, KS에서는 차이가 없었다. 그러므로 이러한 PO시리즈의 특징을 살펴볼 때 투수력과 득점 및 집중력, 수비력을 바탕으로 한 장타력에 주안점을 두고 훈련할 필요성이 있다는 것이다.

KS에서의 경기력의 특징 중 승/패 간의 측정변수에서 나타난 차이의 중요 순위는 세이브〉 타점율〉 방어율〉 WHIP〉 OPS(장타율+출루율)로 나타났으며, 변환된 기술영역변인은 투수력〉 득점 및 집중력〉 수비력〉 타력〉 선구능력 순으로 승/패 간에 유의미한 차이를 보였다. 이러한 특징은 투수력과 수비력을 바탕으로 한 타력과 선구능력의 조화에 집중력을 더한 결과로 여겨진다. 한국시리즈는 최종단계로 경기력이 최고조로 오른 두 팀(승/패)의 경기력의 차이는 미세하여 집중력을 잃은 사소한 실수가 승/패를 가름하는 경우가 있음이 자료 분석에서 나타나고 있다.

두 번째는 각 단기전시리즈에서 구단의 승/패를 예측할 수 있는 예측모형을 산출함에 있어 예측변수의 선택기준에 따라 모형의 차이를 비교하는 것으로 준PO에서의 예측모형비교에 따른 특징은 모든 변수를 예측변수(6개)로 적용한 적중률이 오히려 예측변인이 적게 투입된 통계적 기준(4개)과 전문가 기준(5개)보다 다소 떨어지는 결과가 산출 되었다는 것이다. 그 결과 최종 선택모형은 간명성에 입각하여, 예측변인이 제일 적게 투입되면서 적중률이 가장 높은 인공신경망모형이 최종선택 되었다.

PO에서의 예측모형비교에 따른 특징은 PO에서의 9

가지모형에 의한 각 사례의 적중률을 간명성에 입각하여 살펴본 결과, 전문가 기준에 의해 만든 기술영역변인을 예측변인으로 적용한 인공신경망 모형(전문가 기준:5개)이 다른 두 모형보다는 예측 적중률이 더 좋았다.

KS에서의 예측모형비교에 따른 특징은 모든 변수가 적용되어 예측변인을 만든 인공신경망모형(모든 변수: 6개)이 최종모형으로 선정 되었다. 또한 준PO, PO, KS의 종합적인 적중률을 살펴보면 판별모형 보다는 이항로지스틱회귀모형이, 이항로지스틱회귀모형 보다는 인공신경망모형이 더 우세한 분류정확성을 보이고 있었다.

이러한 연구결과를 토대로 준PO, PO, KS 에 진출한 구단들에게 단기전의 특성을 제시한 이 연구가 사전에 대비 할 수 있는 유용한 정보로 쓰이길 바라며, 여타 다른 종목의 전력평가지수 개발에 긍정적인 효과가 있기를 바란다.

참고문헌

김승대(2003). 야구경기에 있어서 타순별 타격성적이 승패에 미치는 영향 : 프로야구 8개 팀 중심으로. 석사논문. 공주대학교 대학원.

김세형, 강상조, 박재현, 김혜진(2008). 한국프로농구 경기기록 분석에 의한 승패결정요인. 한국체육측정평가학회, 10(1), 1-12.

이대호(2013). "전력분석, 지시한 것 아냐, 사과한다." 2013년 2월 19일 검색, http://osen.mt.co.kr/article/G1109544817

김주학, 노갑택, 박종성, 이원희(2007). 신경망분석을 이용한 축구경기 승·패 예측모형 개발. 한국체육과학회, 18(4), 54-63.

김차용(2001). 프로야구경기 분석을 통한 승·패 예측모형. 한국사회체육학지, 16, 807-819.

박진(1999). Gibbs Sampling을 이용한 한국프로야구에서 상황별 효과분석. 응용통계, 14(12), 121-136.

박승현(2008). 한국프로야구 타자의 고액연봉에 영향을 미치는 경기력 요인. 한국체육학회지, 7(2), 485-494.

배재영, 이진목, 이제영(2012). 주성분회귀분석을 이용한 한국프로야구 순위. 한국통계학회 논문집, 19(3), 367-379.

서재순, 정태충(1993). 프로야구 승·패 예측시스템 개발에 관한 연구. 대한전자공학회 추계종합 학술대회 논문집, 16(2), 93-11.

신상근, 박기철, 조영석, 최세현(2007). 한국프로야구팀의 승패 요인분석에 관한 연구 : 삼성 라이온즈를 중심으로. *The Korean Data Analysis Society, 9*(4), 2071-2083.

승희배, 강기훈(2012). 한국 프로야구 선수들의 경기력과 연봉의 관계 분석. 한국데이터정보과학회지, 23(2), 285-298.

한국야구위원회(2005, 2006, 2010년 2011). 한국야구연감. 서울: 사단법인 한국야구위원회.

한국야구위원회(2009). 한국프로야구기록대백과(1982년-2008년), 제4판.

오광모, 이장택(2003). 데이터마이닝을 이용한 한국프로야구선수들의 연봉에 관한 모형연구. 한국스포츠사회학지, 16(2), 295-309.

이대호(2013. 2. 19). "전력분석, 지시한 것 아냐, 사과한다." http://osen.mt.co.kr/article/G110954817

이장택, 김용태(2005).한국프로야구에 적당한 득점 추정측도에 관한 연구. *Journal of the Korean Date Analysis Society 8*(2), 857-869.

이장택, 김용태(2006a).한국프로야구에서의 승률 추정에 관한 연구. *Journal of the Korean Date Analysis Society 7*(6), 2289-2302.

이장택, 김용태(2006b).한국 프로스포츠에서의 승률 추정. *Journal of the Korean Date Analysis Society 8*(5), 2105-2116.

이영훈(2007). 한국프로야구 경기력 결정요인에 관한 실증분석. 한국체육측정평가학회지, 9(2), 63-77.

이용구, 허준(1999). 데이터 마이닝에서 신경망 분석과 의사결정나무 분석의 비교. 수학·통계논문집 (6).

양병화(2006). 다변량데이타 분석법의 이해. 서울: 커뮤니케이션북스.

조영석, 조용주(2003). 한국 프로야구에서 Beane Count 적용에 관한 연구. *Journal of the Korean Date Analysis Society, 5*(3), 649-658.

조영석, 조용주(2004). 2003시즌 한국 프로야구에서 WHIP가 방어율에 미치는 영향에 관한 연구. *Journal of the Korean Date Analysis Society, 6*(5), 1415-1424.

조영석, 조용주(2005a).한국 프로야구에서 OPS와 득점에 관한 연구. *Journal of the Korean Date Analysis Society, 7*(1), 221-231.

조영석, 조용주(2005b). 한국 프로야구에서 득점과 실점을 이용한 승률 추정에 관한 연구. *Journal of the Korean Date Analysis Society, 7*(6), 2303-2312.

장인식, 원정심(1996). 한국프로야구의 홈경기 효과에 관한 연구. 응용통계, 11(12), 51-70.

정태충(1999). 기계학습 및 시뮬레이션을 이용한 프로야구 경기예측 및 On-line 시뮬레이션 게임 시스템 개발. 경희대학교 산학공동기술개발사업 연 차 연 구 보 고서(정보통신부).

최옥진(2002). 프로야구 선수들의 집단별 성적변화 분석에서

의 평균회귀현상 적용. 교육이론과 실천, 12(2). 407-415
최영근, 김형문(2011). 한국 프로야구 경기결과에 관한 통계적 연구. 한국통계학회, 24(5), 915-930.
최용석, 심희정(1995). 82~92 한국프로야구의 각 팀과 부문별 평균 성적에 대한 추가적 주성분 분석의 응용, 응용통계연구, 8(1), 295-309.
최형준(2009). 2002, 2006년 축구 월드컵 대회를 통한 경기력 분석에 관한 연구. 한국체육측정평가학회, 11(2), 41-52
최형준, 김주학(2006). 인공신경망을 이용한 2005년도 영국 윔블던 테니스 대회의 경기결과 예측에 관한 연구. 한국체육학회지, 45(3), 459-467.
천영진, 최형준(2013). 프로야구 경기에서 상황에 따른 초구 타격의 결과 분석 연구. 한국체육과학회지, 22(1), 1077- 1083.
채진석, 엄한주(2010). 프로야구구단의 성적변화추이와 상대적 전력 비교 평가. 체육과학연구, 21(1), 956-972.
채진석, 조은형, 엄한주(2010). 프로야구 Post시즌 진출 예측을 위한 통계적 모형 비교. 한국체육측정평가학회 12(1), 33-48.
채진석(2011). 프로야구구단의 경기력 비교를 위한 전력평가지수 개발 및 활용. 박사논문. 성균관대학교 대학원.
한국야구위원회(2009). 한국프로야구 기록대백과. 서울시: 사단법인 한국야구위원회.
허명회, 이용구(2003). 데이터 마이닝 모델링과 사례. 서울: (주)데이터 솔루션.
허명회(2008). SPSS데이터 검증, 신경망과 PLS회귀. 서울: (주)데이타솔루션.
허준회, 정태충(1998).프로야구 경기 예측 시뮬레이터에서의 역전파 알고리즘을 이용한 투수 교체시기 예측 모듈 개발. 한국정보과학회 봄 학술발표논문집, 25(1), B
홍석미, 안종일, 정태충(1996). 프로야구 승패 예측을 위한 게임 시뮬레이터 개발에 관한 연구. 한국정보과학회 가을 학술 논문집, 23(2), A
홍세희(2005). 이항 및 다항 로지스틱 회귀분석. 서울: 교육과학사.
홍종선, 최정민(2008). 2007년 한국프로야구에서 도루성공 모형. 응용통계연구, 21(3), 455-468.
황서영(2007). 국내프로야구 선수의 경기력 분석 : KBO기록과 Sabermetrics기록의 차이점을 중심으로. 석사학위논문, 명지대학교.
古俗野 恒, 김명수(2003). 다변량해석 가이드. 서울: 대한미디어.
Agresti, A. (2002). *Categorical Data Analysis*, (2nd ed). New York: John Wiley and Sons.
Hosmer, D. W., & S. Lemeshow. (2000). *Applied Logistic Regression*, (2nd ed). New York: John Wiley and Sons.
Anderson, T. W. (1958). *Introduction to multivariate statistical analysis*. New York: John Wiley & Sons, Inc..
Cooley, W. W., & P. R. Lohnes. (1971). *Multivariate data analysis*. New York: John Wiley & Sons, Inc..
Adler, J. (2006). *Baseball Hacks*. United States of America: O'Reilly Media, Inc.
Costa, Gabriel. B., Huber, Michael. R., & Saccoman, John T. (2008). *Understanding Sabermetrics(An Introduction to the Science of Baseball Statistics*). McFarland & CoIncPub.
Daniel, S. J . (2005). *A regression analysis of predictors on the productivity indices of Major League Baseball: 1985-2003* [Ph.D. dissertation]. United States, Nebraska: The University of Nebraska - Lincoln.
Hosmer, D. W., & S. Lemeshow. (2000). *Applied Logistic Regression*, (2nd ed). New York: John Wiley and Sons.
James, B. (2008). *The Bill James handbook*, New York: Ballantine Books.
Levernier, W & Barilla, A. G. (2006). The probability of winning and the effect of home-field advantage: The case of MLB. *Academy of Information and Management Sciences Journal 9*(2), 61-77.
Michael Lewis. (2003). *Money ball:* W. W. Norton & Company.
Total Baseball(EDT). (2003). *Total Baseball Trivia*, Sportclassic Books, Toronto
Thorn, J. & Palmer, P. (1984). *The Hidden Game of Baseball: A Revolutionary Approach to Baseball and Its Statistics.* New York: Doubleday.

Performance Analysis and a Forecasting Model for The Short-Term Series in The Korean Professional Baseball League

Jin-Seok Chea, & Jong-Kook Song
Kyung Hee University

This study has been conducted to develop methods and techniques for the analysis of data related to baseball performance using the winning and losing games. The purposes of the study were to examine differences of athlete performance for semi playoff, playoff, and Korean professional baseball series and to develop optimal forecasting model for the short term series. Data used in the study were taken from Korean professional baseball association. Three data sets including semi play off from 1982 to 2012, play off from 1989 to 2012, and Korean series from 1982 to 2012 were used. To compare athlete performance by winning and losing games for short-term series t-test was applied. This study created new parameters by weighted value through the equalization process to calculate skill related variables as a predicted variable. Three predicted models such as discriminant, binary logistic regression and artificial neural network models were developed to clarify the suggested models. The results showed that the number of significant parameters increased as the series continued. In particular, a variable related to error was added as a significant variable at the Korean Series. A third base hit in the play-off and a second base hit were also added as significant parameters in the play-off and the Korean series, respectively. In addition, W/L a major variables affecting a given technology area, the pitching PO, PO, the inertia, KS, the pitching, respectively. An artificial neural network model was finally selected with the highest accuracy and lowest input of estimated parameters in the semi play-off. In the play-off, artificial neural network model that applied technical area parameters by specialist criteria had better accuracy rate than two others. In the Korean series, artificial neural network model that created estimation parameters by applying all parameters was chosen as the final model. When the overall accuracy level of semi-play off, play off and Korean series was figured out, binary logistic regression model had higher accuracy of classification than discriminant model, but artificial neural network model had the higher accuracy of classification than binary logistic regression model.

Key Words: Korea Professional Baseball, Post Season, Discriminant Analysis, Logistic Regression, Artificial Neural Network Analysis KISS