

통계모형의 전문스포츠 현장 적용 사례

엄한주 성균관대학교 부교수, 조정환 서울여자대학교 교수,
신승윤 용인대학교 조교수,

요 약

스포츠 통계와 관련된 주제들은 경기결과의 설명 예측력 분석, 선수와 팀 평가, 경기내용의 요약, 경기의 환경적 요인 평가, 경기 규칙 분석, 경기결과의 시각적 표현, 토너먼트 구조 연구 등의 영역으로 다양하다. 이 글에서는 야구 축구 농구 테니스 종목울 중심으로 통계적 방법을 적용하여 스포츠 현장의 문제를 어떻게 접근하고 있는지를 개괄하였다. 전통적으로 스포츠 통계는 선수와 팀 그리고 전략의 평가가 주를 이루었지만, 오늘날 스포츠 통계는 경기의 잠재적, 외적 요인 등의 주제로 관심영역을 넓혀가고 있다. 그러나 국내에서는 경기내용의 기술적(descriptive) 분석이 주류를 이루고 있으며, 전문적 수준의 통계적 접근은 활발하지 못하고 있다. 현장과 자료의 특성을 고려하여 통계와 스포츠(체육)분야의 다양한 협동연구가 필요하다.

주요용어: 스포츠 통계, 선수 팀 평가, 경기력 평가, 득점 모형, 토너먼트 구조, 경기 전략

1. 머리말

대부분의 스포츠는 점수체계(scoring system)를 적용하여 승부를 가린다. 경기 내용의 특성에 따라 시간의 측정, 득점의 셈, 평균 또는 단계적 계산과정 들을 적용하여 스포츠의 질적 양적 특성을 나타내고 그 의미와 가치를 판단하게 된다. 신체적 동작의 미적 기술적 내용을 주관적으로 판정하는 종목(체조, 다이빙, 피겨스케이팅 등)에서도 심판의 판정은 마찬가지로 숫자 형식으로 제시된다. 이러한 배경에서 스포츠 경기 결과로 얻어지는 자료를 통계적 방법으로 분석하고 이해하려는 시도는 다른 분야와 다를 바 없다.

그러나 스포츠의 역사가 상대적으로 오랜 서구에서도 스포츠 현상을 통계적으로 분석하고 이해하려는 노력은 오래되지 않았다고 볼 수 있다. 서구의 경우에서도 70년대에 'Management Sciences in Sports and Optimal Strategies in Sports'가 소개된 이래 관련 연구가 지속되어 오고 있지만 스포츠 통계만을 다루는 전문 학술지 형식으로는 발전되지는 못하였다.

스포츠 현장의 경우에서도 1876년에 미 프로야구 리그가 창설되었지만 프로 야구자료에 대한 통계적 접근은 1960년대에 접어들어 전자계산기, 컴퓨터가 보급되면서 시작되었으며 학계에서는 1992년에 와서 ASA에 스포츠 통계 분과를 개설하게 된다. 국내의 경우는 80년대에 접어들어 소위 수량적 접근 방법에 익숙한 연구자들에 의해 스포츠와 체육 현장의 자료를 통계적 방법으로 분석한 개별 연구결과물들이 보고서 또는 학회지에 소개되는 수준에 있다.

그 동안 스포츠 통계와 관련된 주제들은 경기결과의 설명 예측력 분석, 선수와 팀 평가, 경기내용의 요약, 경기의 환경적 요인 평가, 경기 규칙 분석, 경기결과의 시각적 표현, 토너먼트 구조 연구 등의 영역으로 다양하다. 프로 스포츠 등 대중적으로 발전된 종목들에서는 주로 경기결과의 예측과 관련된 영역들이, 선수나 팀 평가와 같은 기본적 주제는 프로나 아마추어 종목을 막론하고 여러 종목에 걸쳐서 통계적 방법을 적용한 연구들이 소개되었다. 이 글은 그 관심과 요구가 높아지고 있는 스포츠 통계 연구의 지평을 조명하는 관점에서, 그 동안 주요 스포츠 종목을 중심으로 적용되어온 통계적 분석 유형들과 더불어 토너먼트 구조의 통계적 특성을 개

팔하는데 목적이 있다.

2. 선수와 팀의 평가

선수와 팀의 평가는 스포츠 통계의 가장 기본적인 내용이다. 관중 또는 경기 관계자들은 시합 중에 일어나는 상황들을 일일이 또 정확하게 기억할 수 없다. 경기 내용에 대한 통계 정보는 선수와 팀 자체의 이미지를 그대로 반영하게 된다. 선수와 팀 평가를 위한 통계적 절차는 경기내용의 성격에 따라 용이하기도 때로는 매우 어려운 과제일 수 있다. 개방적(open) 연속적 성격의 스포츠 종목에 비하여 폐쇄적(closed) 이산적 성격의 종목에서는 통계적 분석의 틀이 용이할 수 있다.

특히 야구처럼 기회가 균등하게 주어지고 시간적 여유가 허용되는 규칙에서는 선수와 팀별 비교, 공헌도 등의 분석도 용이하고, 수비와 공격도 명확하게 구분되기 때문에 독립적인 비교 분석이 쉽다. 이러한 특성 때문에 야구종목에서는 130여가지가 넘는 다양한 측도(measures)들이 기록 도구로 사용되어, '야구는 완전 복기가 가능한 통계 경기'라는 말로서 회자되기도 한다. 그러나 현장에서는 기술적 자료 즉 공격(타격, 주루플레이)과 수비(피칭, 필드플레이) 영역의 개별 측도들로부터 경기 상황을 고려한 다차원 측도들이 득점력 예측 또는 선수평가 모형으로 소개되고 있다. Quality Starts(선발 6이닝 이상을 던져 3실점 이내로 막은 경기)도 비교적 최근(Dewan 등, 1990)에 소개된 투수 평가 기준이다.

각기 다른 시즌별 성적을 토대로 한 선수 기록의 상대적 비교는 불가능하다. 게임과 출전 횟수, 상대 팀 선수 수준, 기구 장비 특성 등에서 특성이 존재하기 때문이다. 비교를 위한 기록의 표준화(normalization) 방법으로 야구에서는 평균타율, 선두타율, 표준편차를 바탕으로 한 타율의 표준화 방법도 소개되었다(Thorn & Palmer, 1985). Schutz(1995)는 시즌간 특성을 고려하여 평균타율 보다는 시즌간 타율의 안정성을 상대적인 비교 준거로 우수함을 제안하였다.

농구 선수의 플레이 유형을 평가하기 위한 연구(Ghosh & Steckel, 1993)에서는 군집분석 방법을 적용하여 선수들의 플레이 유형을 6가지(Scorers, Dishers, Bangers, The inner court, Walls, Fillers)로 분류하여 각각 특성을 설명하고 대표적인 선수를 예로 들기도 하였다. 이외에도 농구 선수의 평정에 다중회귀분석 방법을 적용한 사례(Bellotti, 1992)에서는 팀 승리에 미치는 긍정적 요인(득점, 공격리바운드, 수비리바운드, 어시스트, 스틸, 브로킹 등)과 부정적 요인(필드골 실패, 프리드로우 실패, 턴 오버), 그리고 파울의 성격 등이 포함되고 있다.

한편 Heeren(1998)은 TENDEX라 칭한 평정 방법을 제안하였는데 위치별 선수의 기여도를 상대적으로 평가하기 위하여 경기 pace 요인을 고려한 표준화 절차를 거쳐 영향력을 분석하였다. Bellotti(1992)도 선수 기여도의 평가의 표준화 방안으로 득점력을 플레이 시간으로 나누어 환산한 점수 산출방법을 제안하기도 하였다.

축구경기의 경우 유럽에서는 정규리그 성적 못지 않게 시뮬레이션 기법을 활용한 축구리그 성적에 팬들의 관심이 집중되고 있다. 시뮬레이션에는 현역선수들의 특성이 그대로 반영될 뿐 아니라 공격과 수비 능력, 홈 그라운드 잇점 요인 등도 포함되고 있다. 그러나 시뮬레이션에 의한 성적은 실제 성적과 대부분 일치하나 부분적으로 차이가 있으며 이러한 결과는 종종 우연적 요소들에 의해 발생된 큰 점수들이 반영되기 때문으로 보고 있다. 퇴장 카드의 효과가 경기결과에 미치는 영향을 포아송 분포를 적용하여 분석한 사례도 보고(Ridder 등, 1994)되고 있다.

3. 환경적 요인 평가

선수, 지도자, 전략, 체력 등 경기 승패에 직접적인 영향을 미치는 요인이외에도 여러 가지 환경적(상황적) 요인들은 직 간접적으로 승패에 영향을 미치게 된다. 관중 효과에서부터 경제적

인 지원, 외적 보상 등 여러 요인들은 무수히 작용할 수 있다. 경기 내적 상황도 시시각각으로 변화하여 설명과 예측을 매우 어렵게 한다. 환경적 요인을 평가하기 위한 통계적 접근도 전형적인 스포츠 통계 주제들이다.

야구에서 팀이 대량(4점 이상) 득점한 다음 투수의 방어율이 다소 높아지는 경향, 투수가 타석에서 안타를 치고 나가는 경우 후속 타자들의 타율이 높아지는 경향, 투수와 타자가 처음 만나는 상황에서 투수가 유리한 경향, 파워형 투수(삼진도 잘 잡지만 볼넷도 많이 허용하는 타입)가 야간 경기에 강한 경향 등에 대한 분석(Dewan 등 1994, Siwoff, 1991, James, 1984) 등은 경기 내적 상황 변화와 경기력 간의 관련성을 다룬 보고들이다. 볼과 스트라이크 조건별 선수 특성별 타율, 좌 우타수 투수별 타율과 방어율, 경기장 효과(사이즈, 모양, 잔디조건, 바람, 습도, 고도, 온도 등) 추정 모형 등도 환경적 요인을 이해하기 위한 지표들이다.

대부분의 스포츠 경기에서 홈 경기의 잇점이 존재한다. 2002년 월드컵에서 한국팀은 홈 경기의 장점을 가장 잘 살린 대표적인 사례로 평가 될 것이다. 축구경기의 홈 경기 잇점은 여러 연구자들에 의해 통계적인 방법으로 소개되고 있는데, 추정의 정확성이 높은 Clarke(1993)의 모형은 다음과 같다. j팀과 홈 경기를 갖는 i 팀의 홈그라운드 잇점(w_{ij})은 $w_{ij} = \mu_i - \mu_j + h_i + \epsilon_{ij}$ 로 나타내고 각 모수(μ_i =i팀의 경기력, h_i =i팀의 홈그라운드 잇점)의 추정을 다음 식으로 제안하였다.

$$H = \text{팀 전체의 득점과 홈 득점간 차이} / (N-1)$$

$$h_i = (\text{전체 팀의 홈 경기 득점과 i팀의 홈 득점 차} - \text{전체 팀의 원정 경기 득점과 i팀의 원정 득점 차} - H) / (N-2)$$

$$\mu_i = (\text{전체 팀의 홈 경기 득점과 i팀의 홈 득점 차} - (N-1) h_i) / N$$

이 모형은 실제 경기 자료와 비교한 결과 관중의 규모, 날씨, 시즌 연도 효과, 팀의 징크스 등에 의해서 다소 차이가 있는 것으로 평가되고 있다.

4. 전략 평가

스포츠는 종목 특성에 따라 선수나 팀의 전술 전략이 매우 정교하게 적용될 수 있는가 하면 적용되기 어려운 경우도 많다. 상대 선수나 팀의 경기력이 비슷하거나 우위에 있을 때 작전의 수행은 더욱 어려워진다. 야구 종목의 경우는 공격과 수비에서 다양한 전략을 매우 정교하게 적용할 수 있는 대표적인 종목으로 여러 통계적인 지표들도 활용되고 있다.

희생타, 도루, 고의 사구, 내외야수 위치 변화, 배팅 오더, 코치의 작전 등 여러 영역의 전략을 평가하기 위한 모형들이 통계적인 자료를 바탕으로 수행되고 있다. 코치의 작전 수행 능력 평가를 위해서 획득점수와 허용점수를 기초로 실제승률과 기대 승률을 산출하고 이를 토대로 D-score를 계산하는 방안(Boyton, 1995)이 제안되기도 하였다.

Bukiet 등(1997)은 마코브 연쇄(Markov chain) 방법을 적용하여 9회 동안 가장 큰 점수를 얻을 수 있는 최적 배팅오더를 구안하였다. 이 연구에서는 적정 라인업을 통해서 시즌통산 30-50점, 6-11승을 기대할 수 있다고 제시하고 적정 라인업에서 소위 장타자가 4번에 가장 약한 타자가 마지막에 포함되지 않는 모형을 제시하였다.

5. 득점 모형

야구종목에서 득점 예측을 위한 모형들은 기본적인 공격 측도들을 조합한 Linear Weight Model(Linsey, 1963), Runs Credit Model(Cook & Garner, 1966), 표준오차 범위를 추가한

Thorn & Palmer(1985)의 득점모형과 도루성공률 모형 등을 들 수 있다. 수비영역에서는 비 에러를 추정을 위한 모형(Neft, 1986), James(1982)의 팀의 승률모형 등이 제안되고 있지만 현장에서의 활용 여부는 알려진 바 없다.

농구의 슛을 통계적 모형으로 접근한 연구도 수 차례 이루어 졌다(Bennett, 1998). 이들 연구는 골의 성공과 실패를 베르누이 시행 모형이나 정상분포(stationary distribution)에 응용하여 확률분포를 여러 관점에서 비교 제시하거나, 슛을 끝인시킨 다음과 실패한 다음의 슛 성공 패턴을 시차 상관관계로 분석하여 우연적 요인에 의한 슛 성공 요인을 설명하고 있다. 이들 연구에서는 특정 선수가 특정한 슛 기회에 연속적으로 골을 성공시킬 수 있는 경우(hot hand)를 시차 자기상관관계로 분석하고 비 독립적인 특성(omnipresent effect) 등으로 설명하고 있다. 프리드로우에서도 시기별 성공률 분포의 비 정상성(non-stationarity), 종속성 특성을 연구한 사례가 소개되고 있다.

축구경기에서는 Pollard 등(1977)과 Maher(1982)이 축구 득점 모형을 포아송(Poisson)분포를 적용하여 제안하였다. 각 팀은 각각 다른 득점력을 가지며 매 경기마다의 득점력 또한 차이가 있다는 전제와, 90분 경기 내에서도 공격과 수비에 치중하는 정도에 따라 득점력이 달라진다는 가정을 적용한 결과 부적 이항분포의 특성을 가지고 있음을 보고하였다. 그리고 패스 횟수(length of pass-move) 또한 같은 분포모양에 적합되고 있음을 제시하였다.

Hill(1974)은 과연 운이 승패를 결정하는지 여부를 검정하기 위하여 프로리그 팀을 대상으로 전문가들의 주관적 판단자료를 근거로 부호화 검정(sign test)을 실시한 결과 기술(skill)이 여러 가지 운의 요소(선수부상, 골대를 맞히는 불운, 심판 판정 등) 보다 승패 결정력이 큰 것으로 나타났지만 경기가 반복될수록 우연적 요소 작용이 줄어드는 경향을 확인하였다.

테니스는 경기의 양상이 상대적으로 단순하고 단절적이어서 경기내용의 기록화 가능성이 높기 때문에 상당수의 수학적 연구들이 이루어져 왔다. 초기의 연구는 이론적 연구가 주류를 이루었는데 이들은 논쟁 가능한 가정들에 기초하여 이루어졌었다. 그러나 최근에는 컴퓨터를 활용한 자료 기록법이 다양하게 개발됨에 따라 실제자료를 근거로 한 연구들이 이루어지고 있다.

테니스는 서브의 성공여부가 결정적으로 영향을 미치는 경기이다. 특히 프로테니스경기에서는 자신의 서브게임은 반드시 이겨야 하고 만약에 서브게임을 패하게 되면 경기전체를 패하게 되는 것이 거의 일반적이다. 따라서 테니스경기에 대한 연구의 대부분은 서브능력의 분석 연구이다. 테니스경기의 기록분석 연구로는 1960년 Kemeny와 Snell이 단식경기의 모형화를 위해 마코프 연쇄방법을 적용한 것을 시작으로 Hsi와 Burych(1971)이 승패 확률을 연구한 사례 등이 있다.

그러나 이들은 한 선수가 서브에서 점수를 얻을 확률은 항상 일정하다는 가정과 각 득점은 독립적이라는 두 가지의 가정 하에 연구를 수행하였으나 이 두 가정은 모두 현장에서 논란이 계속되었었다. 그 후 Gale(1971), George(1973)는 서브에서의 득점이 경기 전체를 좌우한다는 전제 하에 서브득점확률을 극대화할 수 있는 전략을 개발하고자 하였다. 테니스경기에서는 2회의 서브가 허용되는데 2차 서브는 1차 서브의 실패 시에만 허용되는 것으로, 일반적으로는 1차 서브는 강하게, 이것이 실패하였을 경우에는 약하게 2차 서브를 하게 된다. 대부분의 선수들은 2차 서브라는 보완장치가 있기 때문에 1차 서브는 가능한 한 강하게 서브하려고 노력한다. Gale(1971), George(1973)는 다음과 같은 정의를 통해 서브 득점을 위한 확률모형을 개발하였다.

- P(A)=서버가 득점할 확률(event A)
- P(S)=폴트 되지 않은 강서브(event B)
- P(W)=폴트 되지 않은 약서브(event W)

$P(A | S)$ =서브가 강하고 폴트 되지 않을 경우 서버가 득점할 조건확률
 $P(A | W)$ =서브가 약하고 폴트 되지 않을 경우 서버가 득점할 조건확률
 $P(AS)$ =폴트 되지 않는 강한 서브를 넣고 득점할 확률
 $P(AW)$ =폴트 되지 않는 약한 서브를 넣고 득점할 확률

만약 한 선수가 1차 서브를 강하게 서브하고 폴트가 되어 2차 서브를 약하게 서브하여 서버가 득점할 확률은 $P(A)=P(A | S)P(S)+P(A | W)P(W)[1-P(S)]$. 그러나 경우의 수는 위의 한가지 경우만이 아니라 1·2차서브, 강·약서브의 조합에 따라 4가지 경우가 나온다. 이를 정리하면 다음 표1과 같다.

표1. 서브전략별 득점확률모형

| 전략 | 1차 서브 | 2차 서브 | P(A) |
|----|-------|-------|-----------------------|
| SW | 강서브 | 약서브 | $P(AS)+P(AW)[1-P(S)]$ |
| SS | 강서브 | 강서브 | $P(AS)[2-P(S)]$ |
| WW | 약서브 | 약서브 | $P(AW)[2-P(W)]$ |
| WS | 약서브 | 강서브 | $P(AW)+P(AS)[1-P(W)]$ |

그러나, $P(A | S) \geq P(A | W)$ 인 것이 합리적이므로 전략SW가 전략WS보다 합적이고 효과적인 전략이므로 전략SW는 사용되지 않을 것이다. 그리고, 다음과 같이 가정했을 때 $R=[1+P(S)-P(W)]^{-1}$, $Z=P(AW)/P(AS)$, 경우에 따라 다음과 같은 적정전략을 추정할 수 있다.

- $1 \leq Z < R$ 일 경우; 전략SW가 적정
- $Z \leq 1 < R$ 일 경우; 전략SS가 적정
- $1 \leq R < Z$ 일 경우; 전략WW가 적정

세계수준의 프로테니스대회와 실제자료를 통하여 위의 전략들을 검증한 결과, 개인차는 부분적으로 있으나 전략SW가 모두에게서 우수한 전략으로 나타났다. 그러나, 이것은 단순히 서브만을 고려한 경우이고, 선수의 기타 능력을 고려한다면 반드시 전략SW만이 효과적인 것은 아니다. 즉 발리능력이 우수한 선수는 전략SS가 효과적이고, 랠리능력이 우수한 선수는 전략SW가 가장 효과적인 전략이었다.

6. 경기방식의 구조

얼마 전에 막을 내린 제 14회 부산 아시안게임이 시작되는 첫날부터 참가국들의 메달 집계 현황과 각 종목의 금메달 수상자들은 메스컴 뿐만 아니라 직·간접적으로 경기를 관전하는 모든 사람의 관심의 대상이었다. 아시안 게임에 참가한 각 종목은 경기내용의 특성에 따라 시간, 거리, 점수 등을 기준으로 순위를 결정하는 기록경기(육상, 수영, 사격 등)가 있는가 하면, 마라톤과 같이 모든 참가자가 단 한번의 운동수행에 의한 기록으로 순위가 가려지는 종목도 있다. 그리고 주어진 시간 내에 (time-dependent) 상대방보다 많은 점수를 얻거나 (축구, 농구 등), 또는 지정된 점수를 (score-dependent) 먼저 선취해야 승리할 수 있는 (배구, 탁구 등) 경기가 있는가 하면, 전문가(심판)들의 주관적 기준에 의해 선수들의 미적, 기술적 동작이 점수로 비교, 평가되는 체조나 다이빙 같은 경기도 있다.

종목의 특성에 따라 다소 차이는 있지만, 스포츠경기는 주어진 일정 내에 참가팀(또는 선수)의 수를 고려하여 가장 공정하게 최종승자를 가리기 위해 특정한 구조의 경기방식을 채택하여 경기를 진행한다. 단 한번의 경기로 순위를 결정하는 마라톤과 같은 경기방식을 채택하는 종목을 제외한 대부분의 경기는 두 팀/선수가 승패를 겨루게 되며, 출전 팀/선수의 수에 따라 예선,

본선, 준결승, 결승 등의 연속적인 경기를 거쳐 최종승자를 가리는 경기방식을 채택하고 있다. 이러한 특정한 구조의 스포츠 경기방식을 '토너먼트 구조(tournament structure)'라 칭하며, 경기 방식에 대한 스포츠 통계의 연구관심은 다양한 경기방식의 효율성(efficacy)을 비교하는데 있다. 경기방식의 효율성이란 가장 실력이 있는 팀/선수가 최종승자가 될 가능성, 즉 참가자의 능력이 최종순위에 반영되는 확률을 의미하며, 이러한 최종순위가 그 종목의 특성을 고려할 때 타당성(validity)이 있는가를 평가하는 것이다.

현재 현장에서 적용되고 있는 경기방식은 매우 다양하나 가장 보편적으로 채택되고 있는 경기방식을 간략히 살펴보면 (McGarry & Schutz, 1997);

- 1) Single Knock-out (SKO) - 시합에서 승자는 다음 단계에 진출하고 패자는 경기에서 탈락하는 방법,
- 3) Double Knock-out (DKO) - 패자 부활전이 있는 경기방식으로 승자는 다음 단계로 진출하고 패자는 다른 시합의 패자와 다시 시합하는 방법,
- 5) Round Robin (RR) - 경기에 참가한 모든 팀/선수가 한번씩 경기하는 방법,
- 6) RR-RR - 참가자 전원이 RR방식으로 시합하여 순위를 정하고 상위 몇 팀이 본선에서 다시 RR방식으로 경기하는 방법,
- 7) RR-KO - 참가자 전원이 RR방식으로 시합하여 순위를 정하고 상위 몇 팀이 본선에서 예선 순위를 고려하여 짝을 이루어 KO방식으로 경기하는 방법,
- 8) GRR-KO - 참가 팀/선수를 여러 집단(Group)의 나누어 각 집단에서 RR방식으로 시합하여 상위 몇 팀이 본선에 진출하여 KO방식으로 경기하는 방법.

위에서 언급한 경기방식들은 기본적인 형태이며, 이러한 형식을 바탕으로 참가 팀/선수의 수, 경기 일정, 경기 수준(competitiveness), 참가하는 팀/선수의 전적(seeding) 등을 고려하여 종목의 특성과 경기 운영의 목적에 맞는 다양한 경기방식으로 변형되기도 한다. 스포츠 경기에서 최종승자가 결정되는 데는 경기 기술적 요인뿐만 아니라 다양한 심리적, 사회적, 환경적 요인이 작용한다. 그러나 모든 조건이 동등하다고 가정하고, 또 경기에 참가하는 동일한 팀/선수들이 이러한 다양한 경기방식으로 경기를 했을 경우, 어느 방식이 가장 능력 있는 팀/선수가 최종승자가 될 수 있는 확률이 높은지를 시뮬레이션 방법을 이용하여 비교, 분석하는데 관련분야의 연구자들이 관심을 보이고 있다.

7. 경기결과 자료의 시각적 표현

미디어, 정보 산업, 디자인, 전문 연구분야 등의 분야를 주축으로 자료의 시각적 표현 기법이 발전해 가고 있다. 더불어 스포츠 경기 내용이나 결과를 시각적으로 정리 요약하려는 시도 또한 점차 확산되어 가고 있다. 기원전 고대 국가에서의 스포츠 장면들은 벽화로 남아 그 사적 가치를 다하고 있는 반면 오늘날에는 각종 경기자료를 시각적으로 정리 요약하여 설명, 예측, 평가, 상품화 등을 위한 자료로 활용하고 있다.

일반 통계분야에서 자료의 시각적 표현에 대한 관심이 70년대부터 활발하게 논의(Turkey, 1977; Tufte, 1983; Cleveland, 1985)되었다면 스포츠 현장에서는 오래 전인 미 프로야구의 초기부터 자료들을 시각적으로 표현하려는 노력이 있어 왔으며 소프트웨어의 발전과 더불어 가속화되어 오고 있다. 70년대 Williams 와 Underwood(1971)는 '타격의 과학'에서 투구와 스트라이크 존을 공간상의 화소(pixel) 특성으로 제시하였으며 이 기법은 최근까지도 스포츠 전문 채널에서 선수들의 배팅경향(hot/cold zone)을 분석하는데 활용되기도 하였다.

Davenport's(1979)는 '야구의 그래픽'에서 다양한 시각적 표현 방법을 제시하여 특히 T.V. 매체로부터 큰 호응을 받았다고 전해진다. 1997년에 와서는 월드시리즈 경기가 3차원 공간으로

표현되는 기법으로도 발전하였으며, 이러한 경기 내용의 시각적 표현은 표, 그래프, 산점도, 그림(그래픽) 등으로 다양하게 적용되고 있다.

스포츠 분야에서 시각적 자료 표현의 주 의도는 경기내용을 유형별로 분류하거나 요약 정리 하는데 있다. 표로 제시되는 자료의 경우 행과 열별로 경기 내용을 분류하는 틀을 바탕으로 하며, 그래프에서는 점들의 분포, 선, 면적 등으로 자료를 표현한다. Cleveland는 그래프 자료의 구성 원칙을 시각적 선명도(clear vision), 이해 용이성(clear understanding), 척도의 적절성(scale), 보편성(general strategy)으로 제시하였다.

그동안 스포츠 현장에서 적용된 시각적 분석자료 사례들은 미디어 분야에 활용되거나, 선수 선발 및 평가, 기술분석 등 다양한 용도로 쓰이고 있다. 야구에서 스트라이크 존 내의 투구방향 별 평균타율(11×7 행렬) 자료와 히트수별 투구 정보, 산포도를 활용한 US Open(골프) 기록 변화 추이, 년도별 야구 타수 분석, 줄기 및 그림을 활용한 종목별 상해 특성 분석, 군집분석 자료를 시각적으로 제시한 스포츠 특성분석 등을 들 수 있다.

특히 스포츠 기록 자료는 그 특성상 시계열적 의미가 많아 시간대별 기록의 유지, 차이의 변화 패턴은 가장 기본적인 관심사 중의 하나이다. 축구의 실시간 공격주도율과 득점간 관련성(채재성 등 2001), 외야수의 홈런과 도루 기록의 시계열적 분석(90년 동안의 기록)을 통하여 외야수의 기능이 power position에서 speed position으로 변화하는 경향을 분석한 자료(Siwoff 등 1993)는 대표적인 사례이다. 스포츠 현장의 자료를 시각적으로 표현하는 경향은 최근 멀티미디어 소프트웨어와 접목되어 상업적(게임 등)으로도 급속도로 발전되고 있다.

8. 맺음말

스포츠 통계가 '스포츠를 재미있게' 하려는데 있다고 하면 통계분석가들은 어떻게 생각할까? 오늘날 스포츠의 역할과 기능이 다양해지고 있지만, 참여와 관람에 '즐거움'이 그 핵심에 있다. 주 관심 목표가 국가간 경쟁과 한계 도전 또는 마케팅과 오락 등 어디에 있던 그 내면의 질서를 심층적으로 이해하려는 노력은 지속되고 있다. 야구경기에서 관중들의 관심이 줄어든 원인을 투수의 지나친 주자 경계 요인으로 분석한 사례(Siwoff 등 1993)는 현장의 문제를 과학적 방법으로 접근한 좋은 사례일 것이다.

그라운드에서는 징크스라고 불리는 많은 에피소드들을 들을 수 있으며, 과학적 분석 결과를 의심쩍게 보는 시각도 없지 않다. 어느 현장도 마찬가지로 그라운드에는 아직 많은 일들이 기다리고 있다. 무한성에 대한 인간의 도전만큼이나 무궁무진한 스토리들도 끝이 없다. 스포츠의 세계를 보다 객관적인 틀 속에서 이해하려는 노력은 그 자체로서도 충분한 의미가 있을 뿐 아니라 그 즐거움 또한 클 것이다. 덧붙여 구미에서 스포츠 통계 분야 연구에 참여한 통계학자들은 하나 같이 스포츠 현장의 자료를 통계교육에 적극 활용할 것을 역설한다. 학생들이 너무 재미있어 한다는 것이다.

2006년 월드컵에서 대한민국 대표팀은 어떤 성적을 거둘 수 있을까? 우리가 경험한 바와 같이 아이스하키나 야구 종목에 비해서 농구와 축구의 홈 경기 잇점은 크게 작용한다고 알려져 있다. 개방기술 특성을 가진 대부분의 구기 경기에서는 목표물인 공을 잘 다루면서도 나머지 선수들과의 상호작용이 필수적이다. 스포츠의 특성에 대한 바른 이해가 선행되어야만 적정수준의 기대와 건전한 비판도 가능할 것이다.

그동안 종목의 특성을 바탕으로 한 다양한 스포츠 통계 연구가 소개되고 오고 있지만 국내의 경우는 기술적 연구 주제들이 주류를 이룬다. 이번 학술발표회를 계기로 스포츠(체육)와 통계학 분야의 학제간 연구교류가 활성화되고, 이러한 노력으로 우리나라가 진정한 스포츠 선진국으로서 토대를 닦는 계기가 되길 바란다.

참고문헌

- 이성철, 채재성, 민대기(2001) 축구경기력 비교를 위한 인덱스 개발 및 적용에 관한 연구, *한국 사회체육학회지*, 16, 1329-1340.
- Belloti, B.(1992) *The Points Created Basketball book*, 1991-1992. New Brunswick, New Jersey; Night Work Publishing.
- Bennett, J.(Ed)(1998) *Statistics in Sport*, Arnold;London.
- Boynnton, B.(1995) Managers and close games. *The Baseball Research Journal*, 24, 81-87.
- Bukiet, B., Harold, E. R. and Palacios, J. L.(1997) A Markov chain approach to baseball. *Operational Research*, 45(1), 14-23.
- Clarke, S. R.(1993) Computer forecating of Australian rules football for daily newspaper. *Journal of the Operational Research Society*, 44, 753-759.
- Cleveland, W. S.(1985) *The Elements of Graphics Data*. Monterey, CA: Wadsworth.
- Cook, E. and Garner, W. R.(1966) *Percentage Baseball*, Second Edition, Cambridge, MA: the MIT Press.
- Davenport, J. W.(1979) *Baseball Graphics*. Madison, WI: First Impressions.
- Dewan, J., Zminda, D. and STATS, In c.(1990) *The STATS Baseball Scoreboard*. New York: Ballantine Books.
- Dewan, J., Zminda, D. and STATS, In c.(1994) *STATS 1994 Baseball Scoreboard*. New York: Ballantine Books.
- Gale, D.(1971) Optimal strategy for serving in tennis. *Mathematics Magazine*, 5, 197-199.
- George, S. L. (1973) Optimal strategy in tennis. *Journal of the Royal Statistical Society Series C*, 22, 97-104.
- Ghosh, A. and Steckel, J.(1993) Roles in the NBA: there's always room for a big man, but his role has changed. *Interfaces*, 23(4), 43-55.
- Hill, I. D.(1974) Association Football and statistical inference. *Applied Statistics*, 23, 203-208.
- Hsi, B. P. and Burych, D. M.(1971) Games of two players. *Journal of the Royal Statistical Society Series C*, 20, 86-92.
- James, B.(1982) *The Bill James Baseball Abstract 1982*. New York: Ballantine Books.
- James, B.(1984) *The Bill James Baseball Abstract 1984*. New York: Ballantine Books.
- Kemeny, J. G. and Snell, J. L.(1960) *Finite Markov Chains*. Princeton: Van Nostrand.
- Linsey, G. R.(1963) An investigation of strategies in baseball. *Operational Research*, 11, 477-501.
- McGarry, T., & Schutz, R. W.(1997) Efficacy of traditional sports tournament structures. *Journal of the Operational Research*
- Maher, M. J.(1982) Medelling association football scores. *Statistica Neerlandica*, 36, 109-118.
- Neft, D.(1986) Is Ozzie Smith worth \$2,000,000 a season? *The Baseball Research Journal*, 15, 43-48.
- Pollard, R., Benjamin, P. and Reep, D.(1977) *Sport and the negative binomial distribution*. In S. P. Landy and R. E. Machol (eds) *Optimal strategies in Sports*. New York: North-Holland, 188-195.

- Ridder, G., Cramer, J. S. and Hopstaken, P.(1994) Down to ten: estimating the effect of a red card in soccer. *Journal of the American Statistical Association*, 89, 1124-1127.
- Schutz, R. W.(1995) The Stability of individual performance in baseball; an examination of four 5-year periods, 1928-1932, 1948-1952, 1968-1972, 1988-1992. In 1995 Proceedings of the Section on Statistic in Sports. *American Statistical Association*, 39-44.
- Siwoff, S., Hirdt, S., Hirdt, T. and Hirdt, P.(1991) *The 1991 Elias Baseball Analyst*. New York: Fireside Books.
- Siwoff, S., Hirdt, S., Hirdt, T. and Hirdt, P.(1993) *The 1993 Elias Baseball Analyst*. New York: Fireside Books.
- Thorn, J., and Palmer, P.(1985) *The Hidden Game of Baseball*. New York; Doubleday.
- Tufte, E. R.(1983) *The Visual Display of Quantitative Information*. Cheshire: Graphics Press.
- Turkey, J. W.(1977) *Exploratory Data Analysis*. Boston: Addison-Wesley.
- Williams, T. and Uderwood, J.(1971) *The Science of Hitting*. New York: Simon and Schuster.