

## 유럽 리그에서 득점과 실점을 이용한 승점 추정에 관한 연구<sup>†</sup>

신상근<sup>1</sup> · 조용주<sup>2</sup> · 조영석<sup>3</sup>

<sup>1</sup>부산대학교 스포츠과학부 · <sup>2</sup>영남대학교 통계학과 · <sup>3</sup>부산대학교 통계학과

접수 2009년 7월 11일, 수정 2009년 9월 4일, 게재확정 2009년 9월 17일

### 요약

본 연구는 1950년부터 2008년까지의 유럽 5대 리그 축구경기 5170게임의 자료에서 승점이 2점인 시즌과 승점이 3점인 시즌으로 나누어 각 리그의 경기당 득점을 비교하였고, 또한 각국 리그에서 승점이 2점인 시즌과 승점이 3점인 시즌의 경기당 득점 평균을 비교하였다. 경기 당 승점을 추정하기 위하여 경기 당 득점과 경기 당 실점을 독립변수로 하여 추정된 회귀식을 유도하였다. 마지막으로 추정된 회귀식을 한국축구 리그에 적용시켰다.

주요용어: 득점, 승점, 실점, 유럽리그, 한국축구 리그.

### 1. 서론

‘유로 2008’의 스페인 우승은 ‘유로 1984’ 이후 24년만의 결승 진출이었고, ‘유로 1964’ 이후 무려 44년만의 우승이었다. 스페인은 월드컵이나 유로경기에서 항상 우승후보로 평가를 받지만 실제 우승과는 거리가 먼 팀이었다. 항상 결정적인 순간에 벽을 넘지 못하고 번번이 좌절을 하였다. 그러나 팬들은 항상 스페인의 놀라운 테크닉과 물 흐르는 듯한 패스워크에 매료되면서 스페인만의 축구 스타일을 즐겼을 것이다. 이번 우승으로 스페인도 메이저 대회에서 우승할 수 있다는 자신감과 함께 세계축구의 주류로 부상될 것이다.

축구는 야구나 다른 단체 구기 종목들에 비하여 동적인 운동이기 때문에 경기 중 일어나는 상황들을 수치로 표현하기에 매우 어려운 점이 많다. 이러한 문제점 때문에 과학적인 연구가 활발하게 진행되지 못한 것이 현실이다. 사실 축구에서 쉽게 얻을 수 있는 자료는 득점과 실점이다. 야구, 농구, 배구에 비하여 얻을 수 있는 자료가 많이 한정되어 있다. 축구에서의 득점은 야구나 다른 구기 종목에 비하여 단순하다. 상대편의 골대에 공을 넣으면 1득점이 된다. 대부분의 유럽축구 협회는 정규시즌 동안 리그에 소속된 각 클럽이 리그전 방식으로 경기를 갖고, 시즌 결과에 따라 우승팀을 결정하게 된다. 몇몇 상위 팀은 상위 리그로 승격하고 하위 팀들은 하위 리그로 강등되는 승강제가 운영되는 것이 보편적인 리그 운영 방식이다. 또한 많은 나라에서는 해당 축구 협회에서 주관하는 토너먼트 방식의 컵 대회 (FA컵)를 운영하고 있으며, 정규 리그 우승팀과 컵 대회 우승팀 간에 슈퍼컵 대회를 열기도 한다.

유럽 대부분의 리그에서는 ‘승’을 3점, ‘무’를 1점 ‘패’를 0점을 주는 승점제 방식으로 리그 우승을 결정하고 있다. 실제로 유럽의 축구는 단순한 스포츠의 이미지를 넘어 일종의 문화라고 해도 과언이 아

<sup>†</sup> 이 논문은 부산대학교 자유과제 학술연구비 (2년)에 의하여 연구되었음.

<sup>1</sup> (609-735) 부산광역시 금정구 장전동 산30번지, 부산대학교 스포츠과학부, 교수.

<sup>2</sup> (712-749) 경상북도 경산시 대동 214-1, 영남대학교 통계학과, 박사수료.

<sup>3</sup> 교신저자: (609-735) 부산광역시 금정구 장전동 산30번지, 부산대학교 통계학과, 부교수.

E-mail: choys@pusan.ac.kr

니다. 단결력, 협동심을 바탕으로 자국의 축구국가대표팀은 국력을 상징하며 나라마다 고유한 축구스타일을 가지고 있다. 축구자료에 대한 국내연구를 보면 한국 프로축구에서 사용되고 있는 순위 결정기준이 통계적으로 합리적이지 못함을 지적하고 이를 위한 방안을 제시하였다. 모든 팀들의 최종적인 게임수는 동일하겠지만 시즌 중간에는 게임수가 다른 경우가 발생할 수 있다. 이리함에도 불구하고 프로축구의 순위는 총 승점으로 중간 순위를 발표하고 있다. 그래서 총 승점이 아닌 평균 승점을 제안하였다(김혁주, 2004). 한국프로축구의 승률 추정에 관한 연구를 살펴보면, 축구는 무승부가 빈번하게 발생하고 다른 종목에서의 무승부보다 의미가 강하기 때문에, 한국프로축구연맹의 승률계산방식에 따른 방식은 다음과 같이 정의된다.

$$\text{승률} = \frac{\text{승리횟수} + 0.5 \times \text{무승부횟수}}{\text{게임수}}$$

한국 프로 축구에서 승률을 추정하기 위하여 사용된 방법을 살펴보면 빌 제임스가 제안한 피타고리안 승률을 이용하여 이장택과 김용태(2006)는 다음과 같은 식을 유도하였다.  $RS^*$ 와  $RA^*$ 는 각각 승패가 결정된 경기의 점수와 무승부인 경우의 점수를 더한 득점과 실점이다.

$$\text{승률} = \frac{RS^{*1.378}}{RS^{*1.378} + RA^{*1.378}}$$

또한 그들은 연관된 가우스 방법을 이용하여 한국 프로 축구에서 승률을 추정하였다. 추정된 식은 다음과 같다.  $X$ 는 게임당 득점,  $Y$ 는 게임당 실점이다.

$$P(X > Y) = Norm\left(\frac{\bar{X} - \bar{Y}}{\sqrt{Var(X) + Var(Y) - 2Cov(X, Y)}}\right)$$

프로축구의 순위를 결정하는 방식은 승률이 아니라 승점이다. 그래서 본 연구에서는 흔히 유럽의 5대 빅 리그로 불리워지는 잉글랜드의 프리미어리그, 이탈리아의 세리아 A, 독일의 분데스리가, 프랑스의 리그 앙, 마지막으로 스페인의 프리메라리가의 자료를 바탕으로 경기 당 득점 (scored goal per game, SGPG)과 경기 당 실점 (lossed goal per game, LGPG)을 바탕으로 경기 당 승점 (points per game, PtsG)을 추정하는 식을 만들고 이 추정된 회귀식을 국내리그에 적용시키고자 한다.

## 2. 유럽 5대 리그 축구자료의 통계분석

국제축구역사통계연맹 (instrument formation flight system for helicopters, IFFSH)에서는 매년 세계 축구리그 순위를 발표하고 있다. 2007년에는 잉글랜드의 프리미어리그가 1위를 하였고, 다음으로 스페인의 프리메라리가, 이탈리아의 세리아 A, 브라질 세리아 A, 아르헨티나 프리메라 디비시온, 독일의 분데스리가, 프랑스의 리그 앙 순서로 각국의 리그를 평가하였고, 2008년은 아르헨티나의 프리메라 디비시온이 1위를 하였고, 다음으로 브라질 세리아 A, 잉글랜드의 프리미어리그 그리고 스페인의 프리메라리가 순으로 나타났다. 프랑스의 73~74, 74~75, 75~76 세 시즌의 승점제는 다른 프리미어리그와 다르게 독특하였다. 승점 2점을 기본으로 하여 3골차 이상으로 승리를 하면 승점 1점을 보너스로 주었다. 그리고 이탈리아의 05~06 시즌에 승부 조작 사건으로 인하여 06~07 시즌 몇 개의 팀이 승점을 감점당한 상태에서 리그를 진행하였다. 'Siena' 1점, 'Lazio' 3점, 'AC Milan' 8점, 'Reggina' 11점, 'Fiorentina' 15점을 감점 당했다. 독일의 경우는 99~00시즌에 'Eintracht Frankfurt'가 승점 2점, 03~04시즌에는 'FC Kaiserslautern'가 승점 3점을 감점 당했다. 본 연구에서는 유럽의 5대 리그에서 위의 예외적인 경우 67개의 데이터를 제외하고 1950년 이후의 자료를 이용하여 분석하였다.

## 2.1. 경기 당 득점 비교

월드컵에서 승점 3점 제도가 도입된 시기는 1994년 미국월드컵부터 이다. 승점 3점을 도입한 가장 큰 이유는 좀 더 공격적인 축구를 유도하기 위해서이다. 유럽의 리그별 승점 2점 시점과 승점 3점 시점을 살펴보면, 잉글랜드는 81~82시즌인데 잉글랜드는 92~93시즌부터 기존의 축구리그인 풋볼리그와는 독립된 프리미어리그를 운영하였다. 그런데 잉글랜드 전체의 축구 리그역사에서 살펴보면, 풋볼리그의 최상위리그가 프리미어리그로 변한 것뿐이어서 본 연구에서는 풋볼리그와 프리미어리그를 따로 구분하지 않는다.

이탈리아는 94~95시즌, 스페인과 독일은 95~96시즌, 마지막으로 프랑스는 88~89시즌에 처음 승점 3점을 도입하였다가 94~95시즌에 다시 승점 3점을 도입하였다. 승점 2점 일 때의 각 리그별 SGPG와 승점 3점 일 때의 각 리그별 SGPG 평균을 비교한 결과는 표 2.1과 같다. 먼저 승점 2점인 시즌은 유의수준 1%에서 통계적으로 유의한 차이가 있었다. 던컨 방법으로 사후분석 결과 독일이 가장 높았고, 잉글랜드, 그리고 프랑스와 스페인은 같은 그룹이었고, 마지막으로 이탈리아 순서로 SGPG가 높은 것으로 나타났다. 승점 3점인 시즌도 역시 유의수준 1%에서 통계적으로 유의한 차이를 나타내었고, 승점 2점과 마찬가지로 독일이 가장 높았다. 그리고 이탈리아, 잉글랜드, 스페인이 다음 그룹이었고, 프랑스가 가장 낮은 것으로 나타났다.

표 2.1 승점 유형별 리그에 따른 SGPG 비교 (\*p<.05, \*\*p<.01)

승점유형	리그	N	평균	표준편차	F	사후분석
승점2점	잉글랜드	682	1.5145	.35790	113.348**	독일 > 잉글랜드 > 프랑스 = 스페인 > 이탈리아
	이탈리아	752	1.1605	.39293		
	스페인	784	1.3824	.47805		
	독일	574	1.5961	.40615		
	프랑스	776	1.4221	.37734		
	합계	3568	1.4039	.43070		
승점3점	잉글랜드	561	1.3107	.32130	21.998**	독일 > 이탈리아 = 잉글랜드 = 스페인 > 프랑스
	이탈리아	255	1.3092	.35594		
	스페인	264	1.3205	.33006		
	독일	232	1.4276	.34813		
	프랑스	290	1.1646	.27537		
	합계	1602	1.3025	.33328		

승점 2점인 시즌과 SGPG와 승점 3점인 시즌의 SGPG를 리그별로 분석한 결과는 표 2.2와 같다. 모든 리그에서 통계적으로 유의한 차이가 나타났고, 잉글랜드, 스페인, 독일, 프랑스 리그는 승점 2점인 시즌의 SGPG가 승점 3점인 시즌보다 더 높았다. 특히 프랑스와 잉글랜드리그가 차이가 큰 것으로 나타났다. 그러나 이탈리아리그는 승점이 2점인 시즌이 승점이 3점인 시즌보다 SGPG 평균이 높은 것으로 나타났다. 전반적으로 승점 3점을 도입함으로써 공격적인 부분을 강화하고자 하였지만 패했을 때 상대 팀이 얻는 승점이 3점이기 때문에 승점 2점일 때 보다는 오히려 더 SGPG가 낮아지는 결과를 낳았다.

## 2.2. 상관분석

승점을 예측하는 모형을 추정하기 위해서 현재의 모든 리그에서 적용되고 있는 승점 3점을 기준으로 하여 1950년 이후 승점 3점인 시즌의 데이터를 추출하였다. 먼저 각 리그별로 PtsG와 SGPG, SGPG-LGPG 그리고 SGPG/LGPG와의 상관분석을 한 결과는 표 2.3과 같다. 모든 리그에서 PtsG와 SGPG, SGPG-LGPG, SGPG/LGPG는 양의 상관관계를 나타내었고 PtsG와 LGPG는 음의 상관관계를 나타내었다. 모든 리그에서 PtsG와 SGPG-LGPG가 가장 강한 상관관계를 나타내고 있었다.

표 2.2 리그별 승점 유형에 따른 SGPG 비교 (\*p&lt;.05, \*\*p&lt;.01)

리그	승점유형	N	평균	표준편차	t
잉글랜드	2점	682	1.5145	.35790	10.460**
	3점	561	1.3107	.32130	
이탈리아	2점	752	1.1605	.39293	-5.343**
	3점	255	1.3092	.35594	
스페인	2점	784	1.3824	.47805	2.334*
	3점	264	1.3205	.33006	
독일	2점	574	1.5961	.40615	5.920**
	3점	232	1.4276	.34813	
프랑스	2점	776	1.4221	.37734	12.210**
	3점	290	1.1646	.27537	

표 2.3 PtsG와 SGPG, LGPG, SGPG-LGPG, SGPG/LGPG 상관분석 (\*p&lt;.05, \*\*p&lt;.01)

리그	SGPG	LGPG	SGPG-LGPG	SGPG/LGPG
잉글랜드	.844**	-.809**	.954**	.897**
이탈리아	.853**	-.809**	.962**	.920**
스페인	.821**	-.738**	.945**	.919**
독일	.831**	-.769**	.948**	.913**
프랑스	.799**	-.769**	.947**	.927**

### 2.3. 회귀분석

각 리그별로 SGPG와 LGPG를 독립변수로 하여 PtsG를 추정한 회귀식은 표 2.4와 같다. 각 리그별로 거의 비슷한 경향을 보이고 있다. 모든 리그에서  $R^2$  값이 0.89 이상으로 상당히 높게 나타났으며, 잉글랜드, 이탈리아, 스페인, 독일 리그는 PtsG에 SGPG가 LGPG에 비하여 상대적으로 많은 영향을 미치고 있었고, 프랑스 리그는 LGPG가 SGPG에 비하여 상대적으로 많은 영향을 미치고 있었다.

표 2.4 SGPG와 LGPG를 이용한 회귀분석 (\*p&lt;.05, \*\*p&lt;.01)

리그	적합된 회귀식	F	R <sup>2</sup>	adj - R <sup>2</sup>
잉글랜드	$\hat{y}_1 = 1.299 + 0.668 \times \text{SGPG} - 0.617 \times \text{LGPG}$	2841.270**	.911	.910
이탈리아	$\hat{y}_1 = 1.295 + 0.709 \times \text{SGPG} - 0.665 \times \text{LGPG}$	1451.419**	.925	.924
스페인	$\hat{y}_1 = 1.349 + 0.649 \times \text{SGPG} - 0.635 \times \text{LGPG}$	1082.483**	.892	.892
독일	$\hat{y}_1 = 1.284 + 0.643 \times \text{SGPG} - 0.584 \times \text{LGPG}$	1019.456**	.899	.898
프랑스	$\hat{y}_1 = 1.361 + 0.693 \times \text{SGPG} - 0.699 \times \text{LGPG}$	1247.401**	.897	.896

각 리그별로 SGPG/LGPG를 독립변수로 하여 PtsG를 추정한 회귀식은 표 2.5와 같다. 각 리그별로 거의 비슷한 경향을 보이고, 모든 리그에서  $R^2$  값이 0.80 이상으로 상당히 높게 나타났다.

표 2.5 SGPG/LGPG를 이용한 회귀분석 (\*p&lt;.05, \*\*p&lt;.01)

리그	적합된 회귀식	F	R <sup>2</sup>	adj - R <sup>2</sup>
잉글랜드	$\hat{y}_2 = 0.713 + 0.591 \times \text{SGPG/LGPG}$	2310.487**	.805	.805
이탈리아	$\hat{y}_2 = 0.589 + 0.685 \times \text{SGPG/LGPG}$	1385.597**	.846	.845
스페인	$\hat{y}_2 = 0.593 + 0.726 \times \text{SGPG/LGPG}$	1428.857**	.845	.844
독일	$\hat{y}_2 = 0.622 + 0.688 \times \text{SGPG/LGPG}$	1154.852**	.834	.833
프랑스	$\hat{y}_2 = 0.597 + 0.707 \times \text{SGPG/LGPG}$	1759.202**	.859	.859

표 2.6 SGPG-LGPG를 이용한 회귀분석 (\*p&lt;.05, \*\*p&lt;.01)

리그	적합된 회귀식	F	R <sup>2</sup>	adj - R <sup>2</sup>
잉글랜드	$\hat{y}_3 = 1.366 + 0.644 \times (\text{SGPG} - \text{LGPG})$	5660.133**	.910	.910
이탈리아	$\hat{y}_3 = 1.352 + 0.688 \times (\text{SGPG} - \text{LGPG})$	3107.343**	.925	.924
스페인	$\hat{y}_3 = 1.367 + 0.643 \times (\text{SGPG} - \text{LGPG})$	2172.172**	.892	.892
독일	$\hat{y}_3 = 1.369 + 0.616 \times (\text{SGPG} - \text{LGPG})$	2029.426**	.898	.898
프랑스	$\hat{y}_3 = 1.354 + 0.696 \times (\text{SGPG} - \text{LGPG})$	2503.291**	.897	.896

각 리그별로 SGPG-LGPG를 독립변수로 하여 PtsG를 추정된 회귀식은 표 2.6과 같다. 각 리그별로의 비슷한 경향을 보이고, 모든 리그에서 R<sup>2</sup> 값이 0.88 이상으로 상당히 높게 나타났다.

추정된 모형의 적합성을 살펴보면 먼저 로지스틱 회귀분석이나 의사결정나무 모형에서 예측모형의 적합성을 알기 위하여 정분류율과 점근적 표준오차 제공근 (root asymptotic square error, root ASE)을 이용한다. 좋은 모형일수록 정분류율이 높아지고, 점근적 표준오차 제공근은 0에 가까워진다 (Choi 등, 2008).

선형 회귀모형확인 방법의 가장 좋은 방법은 새로운 자료를 수집하여 선택된 모형을 검토하는 것이다. 새로운 자료에서 설명변수들의 특정한 값을 모형에 대입했을 때의 적합치가 새로운 자료의 반응치와 얼마나 차이가 나는지를 보기 위하여 모형에 대한 평가 척도를 위하여 평균제곱오차 (mean square error, MSE), 평균절대편차 (mean absolute deviation, MAD) 그리고 평균절대백분율오차 (mean absolute percent error, MAPE) 등을 활용한다. MAPE는 수요의 크기에 상대적인 예측오차를 측정하며, 예측 성과를 비율의 관점에서 측정한다. MAPE는 다음과 같이 정의한다. 여기서,  $y_i$ 는 새로운 자료의 관측값,  $\hat{y}_i$ 는 최종모형을 이용한 예측값,  $n$ 는 자료의 개수를 말한다 (박상규와 오정현, 2009).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100.$$

본 연구에서는 일반적으로 많이 사용되는 추정량 선택기준인 MSE와 MAD를 사용하였다. 어떤 추정량이 더 좋은가에 대한 평가는 각 추정량의 MSE와 MAD를 구해서 구한 값이 가장 작은 추정량이 제일 좋은 추정량이라 할 수 있다. 여기서,  $y_i$ 는 새로운 자료의 관측값,  $\hat{y}_i$ 는 최종모형을 이용한 예측값,  $n^*$ 는 자료의 개수를 말한다.

$$MSE = \sum_{i=1}^{n^*} \frac{(y_i - \hat{y}_i)^2}{n^*}, \quad MAD = \sum_{i=1}^{n^*} \frac{|y_i - \hat{y}_i|}{n^*}.$$

적합된 회귀식의 MSE와 MAD를 비교하여 보면 표 2.7과 같다. MSE와 MAD를 동시에 고려한 결과를 살펴보면 전반적으로 MSE 측면에서는 프랑스를 제외한 잉글랜드, 스페인, 이탈리아, 독일은 SGPG와 LGPG를 독립변수로 하여 PtsG를 추정된 회귀식이 좋았으며, MAD 측면에서는 잉글랜드, 스페인은 SGPG와 LGPG를 독립변수로 하여 PtsG를 추정된 회귀식이 가장 좋았고, 이탈리아, 독일, 프랑스는 SGPG-LGPG를 독립변수한 회귀식이 가장 좋은 것으로 나타났다. 전반적으로 SGPG와 LGPG를 독립변수로 한 회귀식이 적합성이 좋음을 알 수 있다.

추정된 경기 당 승점과 실제값 사이의 오차를 알아보기 위해 MSE와 MAD를 이용하여 위와 같은 결과를 유도하였는데 SGPG와 LGPG를 독립변수한 회귀식과 SGPG-LGPG를 독립변수한 회귀식이 실제 경기 당 승점과 최솟값, 최댓값 그리고 평균의 정도가 비슷함을 알 수 있다.

표 2.7 추정된 승점에 대한 MSE와 MAD

리그	$\hat{y}_1$		$\hat{y}_2$		$\hat{y}_3$	
	MSE	MAD	MSE	MAD	MSE	MAD
잉글랜드	<b>0.01204</b>	<b>0.08715</b>	0.02621	0.11794	0.01215	0.08748
이탈리아	<b>0.01320</b>	0.09055	0.02733	0.11737	0.01333	<b>0.09054</b>
스페인	<b>0.01224</b>	<b>0.09042</b>	0.01760	0.10132	0.01227	0.09047
독일	<b>0.01336</b>	0.09480	0.02183	0.11450	0.01337	<b>0.09451</b>
프랑스	0.01070	0.08253	0.01448	0.09354	<b>0.01057</b>	<b>0.08211</b>

표 2.8 추정된 승점의 최댓값, 최솟값

리그	구분	경기 당 승점	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$
잉글랜드	최솟값	0.29	0.21	0.85	0.20
	최댓값	2.50	2.38	3.55	2.38
	평균	1.3660	1.3658	1.3659	1.3660
이탈리아	최솟값	0.35	0.33	0.78	0.32
	최댓값	2.55	2.34	3.04	2.34
	평균	1.3496	1.3504	1.3501	1.3497
스페인	최솟값	.34	.54	.87	.54
	최댓값	2.24	2.20	2.50	2.19
	평균	1.3670	1.3672	1.3665	1.3667
독일	최솟값	.53	.56	.90	.55
	최댓값	2.29	2.24	2.85	2.24
	평균	1.3708	1.3702	1.3700	1.3711
프랑스	최솟값	.45	.61	.90	.60
	최댓값	2.21	2.13	2.40	2.12
	평균	1.3542	1.3615	1.3536	1.3545

#### 2.4. K-리그 적용

유럽 5대 리그 자료를 바탕으로 K-리그에 회귀식을 적용한 결과는 표 2.9와 같다. K-리그에 적용시킨 결과는 유럽의 5대리그를 적용시킨 결과와 비교하여보면, 프랑스 리그와 적합성이 일치함을 알 수 있다. 물론 이 결과로 K-리그와 프랑스 리그가 성격이 유사하다고 판단하기는 어렵다.

표 2.9 K-리그에서 추정된 승점에 대한 MSE와 MAD

K-리그	$\hat{y}_1$		$\hat{y}_2$		$\hat{y}_3$	
	MSE	MAD	MSE	MAD	MSE	MAD
2007시즌	0.01617	0.10613	0.03262	0.14311	<b>0.01589</b>	<b>0.10435</b>
2008시즌	0.08995	0.27779	0.27853	0.50651	<b>0.08603</b>	<b>0.27140</b>

### 3. 결론

축구는 전술과 전략 및 선수들의 개인기, 팀 전체의 포메이션 등 많은 부분들에 대하여 논의를 할 수 있으나 축구도 승패를 결정하는 스포츠인 관계로 역시 승점을 얻는데 중요한 요소는 득점이다. 이러한 팀의 SGPG와 LGPG를 바탕으로 PtsG를 추정하여 본 결과 잉글랜드, 이탈리아, 스페인, 독일, 프랑스 각 리그의 고유한 축구 스타일이 있고 흐름이 있는 것은 사실이지만 승점에 미치는 득점과 실점의 영향을 거의 차이가 없었다.

승점이 2점인 시즌과 승점이 3점인 시즌을 각 리그별로 SGPG를 비교한 결과는 일반적으로 잉글랜드

와 스페인이 공격적이라는 견해와는 다르게 분석되었다. 승점이 2점인 시즌과 3점인 시즌 모두 독일의 SGPG가 가장 높았다. 그리고 각 리그별로 승점이 2점인 시즌과 3점인 시즌의 SGPG를 비교한 결과 이탈리아 리그를 제외하고는 2점일 때의 SGPG가 더 높은 것으로 나타났다.

SGPG와 LGPG를 이용하여 회귀분석을 한 결과를 보면, 잉글랜드, 이탈리아, 독일, 스페인 리그에서는 SGPG와 LGPG를 독립변수로 한 회귀식의 MSE가 가장 좋은 것으로 추정되며, MAD 측면에서는 프랑스, 독일, 이탈리아 리그는 SGPG-LGPG를 독립변수로 한 회귀식이 가장 좋았다.

마지막으로 K-리그 2007년, 2008년 자료에 적용시켜 본 결과는 MSE, MAD 모든 측면에서 SGPG-LGPG를 독립변수로 한 회귀식의 적합도가 더 좋았다. 유럽의 5대 리그에서 추정된 회귀식을 K-리그에 적용시켰을 때, 프랑스리그와 비슷한 결과를 보여 주었다. 앞으로 좀 더 다양한 자료를 추가하여 K-리그에 맞는 승점을 추정하는 새로운 모형의 개발이 필요하다고 생각한다.

### 참고문헌

- 김혁주 (2004). 한국 프로야구와 프로축구의 순위 결정기준은 통계적 관점에서 합리적인가? <한국자료분석학회지>, **6**, 1767-1775.
- 박상규, 오정현 (2009). 신제품 수요예측을 위하여 누적자료를 활용한 회귀모형에 관한 연구. <한국데이터정보과학회지>, **20**, 117-124.
- 이장택, 김용태 (2006). 한국 프로스포츠에서의 승률 추정. <한국자료분석학회지>, **5**, 2105-2116.
- Choi, S. B., Kang, C. W. and Cho, J. S. (2008). Feature analysis for seceders among new students passed the D university entrance examination. *Journal of the Korean Data & Information Science Society*, **19**, 1111-1122.

# A study on points per game using scored goal per game and loss goal per game in the union of European football professional league<sup>†</sup>

Sang-Keun Shin<sup>1</sup> · Yong-Ju Cho<sup>2</sup> · Young-Seuk Cho<sup>3</sup>

<sup>1</sup>Division of Sport Science, Pusan National University

<sup>2</sup>Department of Statistics, Yeungnam University

<sup>3</sup>Department of Statistics, Pusan National University

Received 11 July 2009, revised 4 September 2009, accepted 17 September 2009

## Abstract

This study used data of soccer match 5170 games from 1950 to 2008 in five European football professional leagues. We compared average of SGPG (scored goal per game) in each two and three points of win. And we compared average of SGPG in each leagues. In order to predict PtsG (points per game), we executed regression analysis using SGPG and LGPG (loss goal per game). Finally, We applied regression analysis to a K-league.

*Keywords:* European football professional league, K-league, LGPG, PtsG, SGPG.

---

<sup>†</sup> This work was supported for two years by Pusan National University Research Grant.

<sup>1</sup> Professor, Division of Sport Science, Pusan National University, Busan 609-735, Korea.

<sup>2</sup> Lecturer, Department of Statistics, Yeungnam University, Gyeongsan 712-749, Korea.

<sup>3</sup> Corresponding author: Associate professor, Department of Statistics, Pusan National University, Busan 609-735, Korea. E-mail: choys@pusan.ac.kr