

한국프로야구 경기 내 역전의 의미 분석

최유정, 강지원, 김혜정, 서유정, 한세미

경기대 응용통계학과

흔히들 '야구는 9회말 2아웃까지는 모른다'라고 한다. 한치 앞을 모르는 것이 야구 경기이므로 승패를 속단하여 포기하지 말라는 뜻이다. 이것이 정말 일리가 있을까? 그 사실확인을 위하여 데이터 분석을 실시하였다. 본 분석은 '역전'에 대한 것이다. 팀 별 역전의 비율과 역전 사실이 해당 경기의 결과에 영향을 미쳤는지를 보았다. 해당 데이터는 2017년부터 2019년까지, 각각의 정규시즌 내 경기 별 점수표를 사용하였다. 이는 웹사이트¹ 크롤링을 통해 얻었다. 분석에서 사용된 '역전'의 정의는 다음과 같다.

역전 7회 이후의 회차에서 승패의 결과가 뒤바뀌는 것. 단, 동점이 되는 경우는 고려하지 않는다.

데이터추출

공식홈페이지(KBO: <https://www.koreabaseball.com/>)에서는 원하는 자료를 제공하지 않는다. 그러므로 웹페이지에 나와있는 HTML코드를 이용, 크롤링을 통해 데이터를 추출하였다. 웹을 통한 단순 반복작업을 처리하기 위해서 python의 selenium²과 chromedriver³를 사용하였다.



순위	팀명	경기	승	무	패	승률	승차	타율
1	SK	62	40	1	21	0.655	6.0	0.254
2	두산	63	40	0	23	0.635	1.0	0.274
3	LG	61	35	0	26	0.574	5.0	0.260
4	기아	64	34	0	30	0.531	7.5	0.262
5	KTC	61	32	0	29	0.525	8.0	0.266
6	삼성	61	28	0	33	0.459	12.0	0.267
7	한화	61	27	0	34	0.443	13.0	0.261
8	KIA	62	25	1	36	0.409	15.0	0.267
9	SSG	61	25	0	36	0.397	16.0	0.270
10	롯데	62	23	0	39	0.371	17.0	0.265

[그림 1] selenium을 이용하여 실행한 브라우저 창

¹ 스포조이(<http://www.spojjoy.com/live/>)

² 웹 기반 관리 작업을 자동화하는 라이브러리

³ 브라우저별로 제공하는 driver. 가상 브라우저 실행 후 여러가지 작업을 수행할 수 있다.

분석결과

다음은 회차별로 역전한 팀과 그 횟수를 추출해낸 결과이다.

1. 회차 별 팀의 역전 횟수

[7회]

KIA	KT	LG	NC	SK	넥센	두산	롯데	키움
6	4	7	8	7	11	5	6	1

[8회]

KIA	KT	LG	NC	SK	넥센	두산	롯데	키움
9	8	5	7	4	5	7	8	6

[9회]

KT	LG	NC	SK	넥센	두산	롯데	삼성	키움	한화
16	22	30	14	6	8	22	16	16	9

[10회]

KIA	KT	LG	NC	SK	넥센	두산	롯데	삼성	한화
6	3	1	5	4	1	1	3	3	4

[11회]

KIA	KT	LG	NC	SK	롯데	한화
1	1	1	2	2	3	4

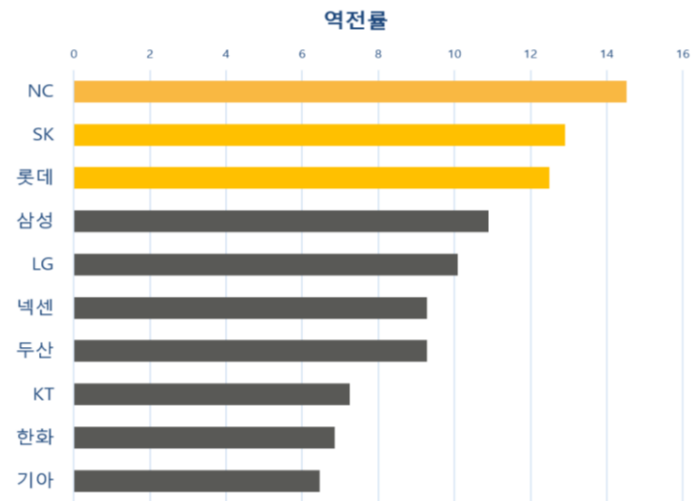
[12회]

LG	NC	넥센	두산	롯데	한화
2	4	2	2	2	2

[표 1] 회차 별 팀의 역전 횟수

보통의 야구경기는 9회에서 끝을 보인다. 그러므로 10회부터는 경기의 수 자체가 적어, 역전을 한 팀과 그 횟수가 급감한다. 데이터 내 전체 경기의 수가 3114개인 것을 감안했을 때 역전이 발생한 경기의 수는 현저히 낮은 비율을 보인다. (약 10%) 이는 7회 이후 역전이 발생할 가능성이 낮다는 것을 나타낸다

2. 역전률(승)



[그림 2] 팀 별 역전률

팀 별로 전체 역전의 경기 중 해당 팀의 역전승을 나타낸 비율 지표이다. NC는 14.5% SK는 13%, 롯데는 12.7%로 역전승을 거둔 상위 3팀에 위치하고 있다. 이를 정규시즌 순위와 비교하여 역전승이 최종 결과에 미치는 영향을 보고자 하였다.

순위	팀	경기수	승	패	무승부	개입차	승률
1	KIA 타이거즈	144	87	56	1	0.0	0.608
2	두산 베어스	144	84	57	3	2.0	0.596
3	롯데 자이언츠	144	80	62	2	6.5	0.563
4	NC 다이노스	144	79	62	3	7.0	0.560
5	SK 와이번스	144	75	68	1	12.0	0.524
6	LG 트윈스	144	69	72	3	17.0	0.489
7	넥센 히어로즈	144	69	73	2	17.5	0.486
8	한화 이글스	144	61	81	2	25.5	0.430
9	삼성 라이온즈	144	55	84	5	30.0	0.396
10	kt wiz	144	50	94	0	37.5	0.347

[그림 3] 2017년 정규시즌 최종 순위표

순위	팀	경기수	승	패	무승부	개입차	승률
1	SK 와이번스 ^[48]	144	78	65	1	-	0.545
2	두산 베어스	144	93	51	0	-14.5	0.646
3	한화 이글스	144	77	67	0	1.5	0.535
4	넥센 히어로즈	144	75	69	0	3.5	0.521
5	KIA 타이거즈	144	70	74	0	8.5	0.4861
6	삼성 라이온즈 ^[49]	144	68	72	4	8.5	0.4857
7	롯데 자이언츠	144	68	74	2	9.5	0.479
8	LG 트윈스	144	68	75	1	10	0.476
9	kt wiz	144	59	82	3	18	0.418
10	NC 다이노스	144	58	85	1	20	0.406

[그림 4] 2018년 정규시즌 최종 순위표

위의 두 정규시즌 최종 순위를 도합하여 도출해낸 2017-2018 종합 순위는 다음과 같다.

팀명	두산	SK	KIA	롯데	넥센	한화	NC	LG	삼성	KT
순위	1	2	2	4	5	5	7	7	9	10

[표 2] 2017-2018 정규시즌 종합 순위표

팀명	NC	SK	롯데	삼성	LG	넥슨	두산	KT	한화	KIA
순위	1	2	3	4	5	6	7	8	9	10

[표 3] 역전률 순위표

정규시즌 종합 1위는 두산, 최하위 성적을 기록한 팀은 KT이다. 하지만 이에 반해 역전률에서는 NC가 1위, KIA가 최하위를 차지하였다. 세부적으로 두 순위를 비교해보았을 때, 동일한 성적을 거둔 팀은 SK 하나 뿐이었고 유사한 성적을 보이는 팀은 롯데와 넥센, 두 팀이었다. 종합 시즌 각각 1위와 3위에 위치한 두산과 KIA는 역전률에서 7위와 10위를 기록하였다. 이를 통해 역전률과 우승팀의 특정 관계는 존재하지 않는다고 볼 수 있다. 나아가 승패의 변동이 일어날 가능성이 낮음을 내포하기도 한다. 결국 '역전'은 야구 경기에서 최종 승리를 이끌기에는 다소 미미한 반전인 것이다.

결론 및 한계점

분석 시행 후 도출된 관계는 '역전률과 승률은 특정 관계를 보이지 않는다'이다. 이는 기대했던 결과가 '역전률과 승률이 양의 관계를 가진다'임에 반하는 것이다. 덧붙여 '역전'이 쉬이 일어나는 일도 아니라고 보여진다. 결국 '야구는 9회말 2아웃까지는 모른다'라는 말은 신빙성이 떨어진다고 볼 수 있다. 본 분석에서 사용된 원 데이터의 경기는 약 3000개이다. 하지만 전처리를 통해 얻어낸 해당 데이터의 경기가 약 300개뿐이었다는것을 감안한다면, 보다 많은 데이터를 사용하여 시행할 시 현재와 다른 결과가 도출될 것으로 예상된다. 또한 종합 순위와 역전률의 순위가 같은 SK가 존재하는 것을 보아 역전률이 승리에 미치는 영향이 아주 없는 것은 아님을 알 수 있다.

참고문헌

신규식 외 1인, 「R 프로그램을 이용한 한국프로야구의 승리 요인과 패턴 분석」, 『2014 대한 산업 공학회』, 한국경영과학회 춘계공동학술대회 논문집

서영진 외 2인, 「기계학습 기법을 이용한 한국프로야구 승패 예측 모델」, 『Journal of The Korea Society of Computer and Information』, Vol.24 No.2, pp.17-24, February 2019

김주희, 「데이터 분석을 위한 프로 야구 기록 예측 모델 및 시각화 도구」, 한양대학교 대학원 석사학위 논문, 2017 02