

한국 축구 국가대표, 최고의 골 결정력을 보여 준 선수는? - Empirical Bayes estimation을 사용한 골 결정력 계산

팀/이름: 현교/ 김현교
소속: NHND&T

Intro.

골 결정력? = 골/슛. 정말 간단한 계산. 그러나, 정말 간단할까요?

Q1.

- 여러분은 이 두 선수의 결정력이 같다고 생각하시나요?

2003년 챔피언스리그 올드트래포드에서...

“3번의 슈트로 해트트릭을 기록한 호나우두”
VS. “1번의 슈트로 1골을 기록한 베컴”

3샷 3킬의 호나우두가 더 뛰어난 골 결정력을 가졌다고 생각하는 이유?

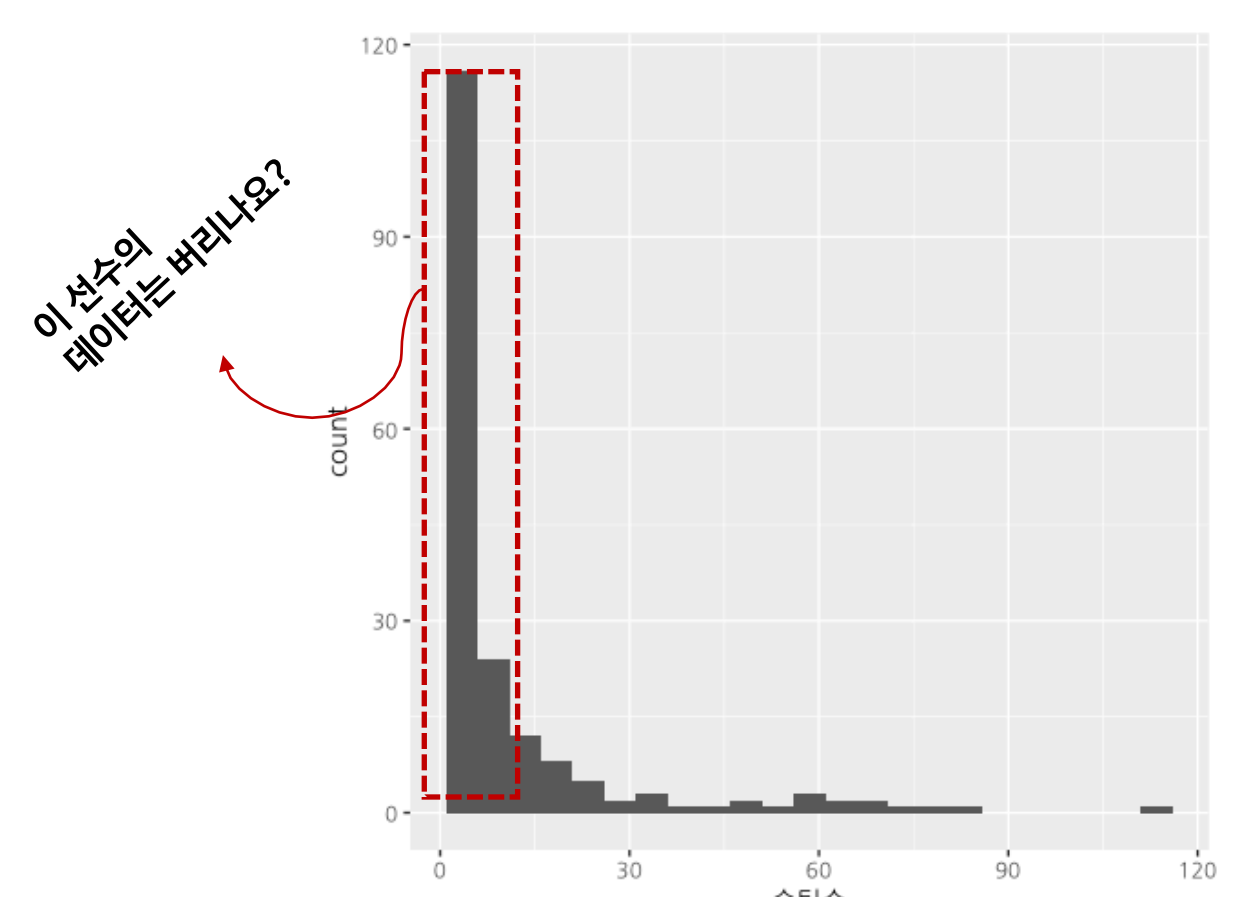
우리가 이미 알고 있는 정보를 기반으로 평가하기 때문. Ex, 1번의 슈트보다 3번의 슈트 100% 골로 이어질 가능성이 더 낮음

Q2.

- 5번 이상 슈트를 한 선수를? 10번 이상 슈트를 한 선수를? 가지고 이야기 해야할
까요? 그렇다면, 데이터의 70%는 버리고 시작.

특히, 국가대표 데이터는 ...

“KFA 데이터 기준, 50%의 선수가 슈팅 5번을 기록하지 못함.
그리고 70%의 선수가 10번을 기록하지 못함.”



1.

Method

“Empirical Bayes estimation”:

전체 데이터로부터 얻은 사전 분포를 개별 데이터에 적용하여 정보를 업데이트하는 방법

Case)

EPL에서,
Player A: 10번 슈팅에 1득점,
Player B: 100번 슈팅에 10득점.

1골/ 10슛
(10%)

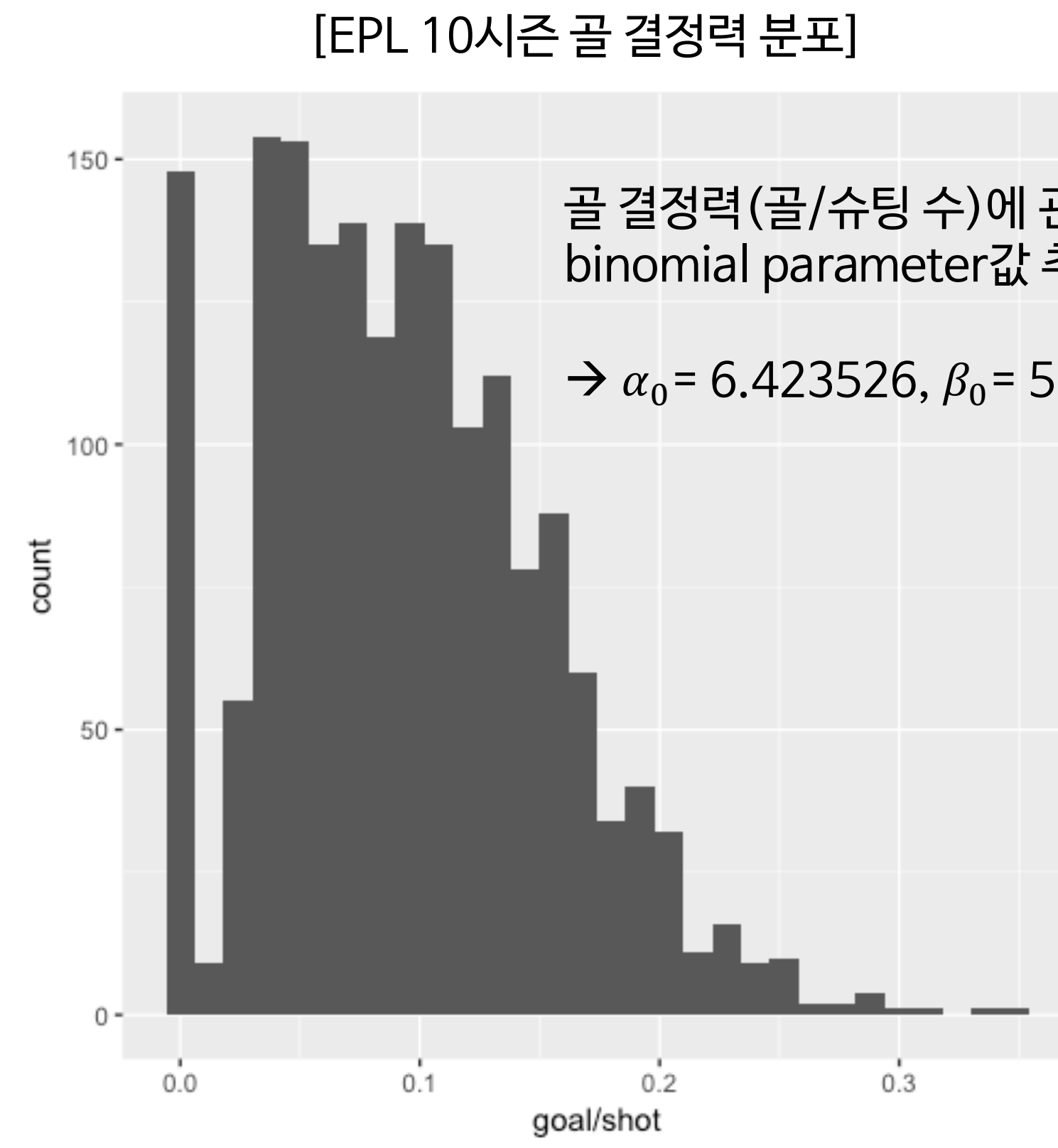


Player A

10골/ 100슛
(10%)



Player B



[EPL 10시즌 골 결정력 분포]

골 결정력(골/슈팅 수)에 관한
binomial parameter값 추정 결과.

→ $\alpha_0 = 6.423526$, $\beta_0 = 59.91984$.

$(goal + \alpha_0) / (shot + \alpha_0 + \beta_0)$

Where,
 $shot$ = the number of shots
 $goal$ = the number of goals

Player A=
 $(1 + 6.423526) / (10 + 6.423526 + 59.91984) =$ 약
0.0972
Player B=
 $(10 + 6.423526) / (100 + 6.423526 + 59.91984) =$ 약
0.0987

Player A < Plyaer B

출처:EPL Official site

2.

Data

기간) 1986년 ~ 2016년, 약 20년간의 남자 성인 국가대표팀 데이터

대상) 301경기의 251명 선수의 이벤트 데이터중 적어도 1번의 슈팅을 기록한 186명의 선수가 분석 대상

* 16골이나 넣은 2003년 네팸 과의 경기는 분석에서 제외

슈팅수	골
슈팅, 위협슈팅, 유효슈팅 성공 수를 슈팅 데이터로 활용 총 2,143번의 슈팅	골인 데이터를 득점 데이터로 활용 총 461골

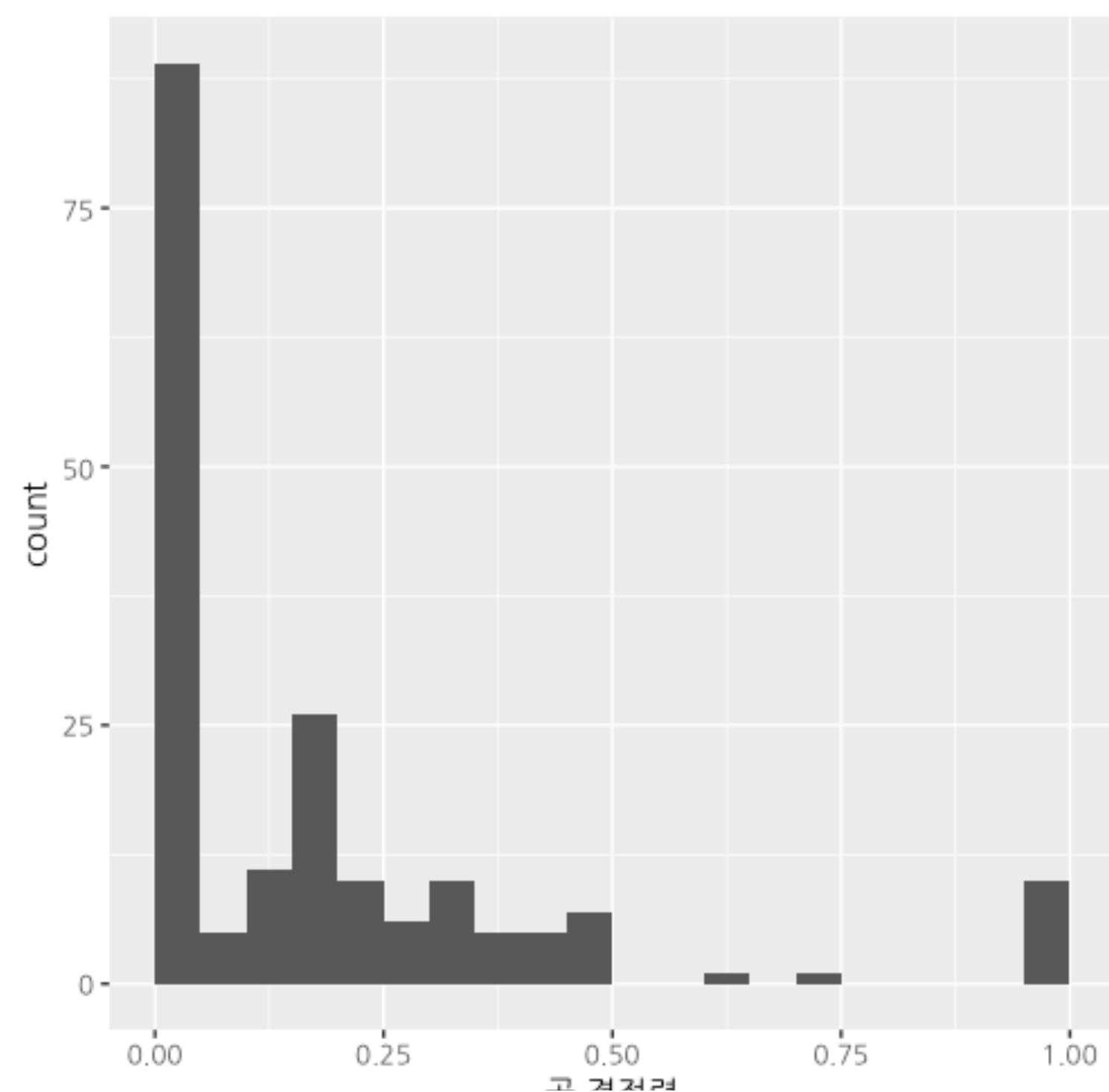
3.

국가대표의 레벨

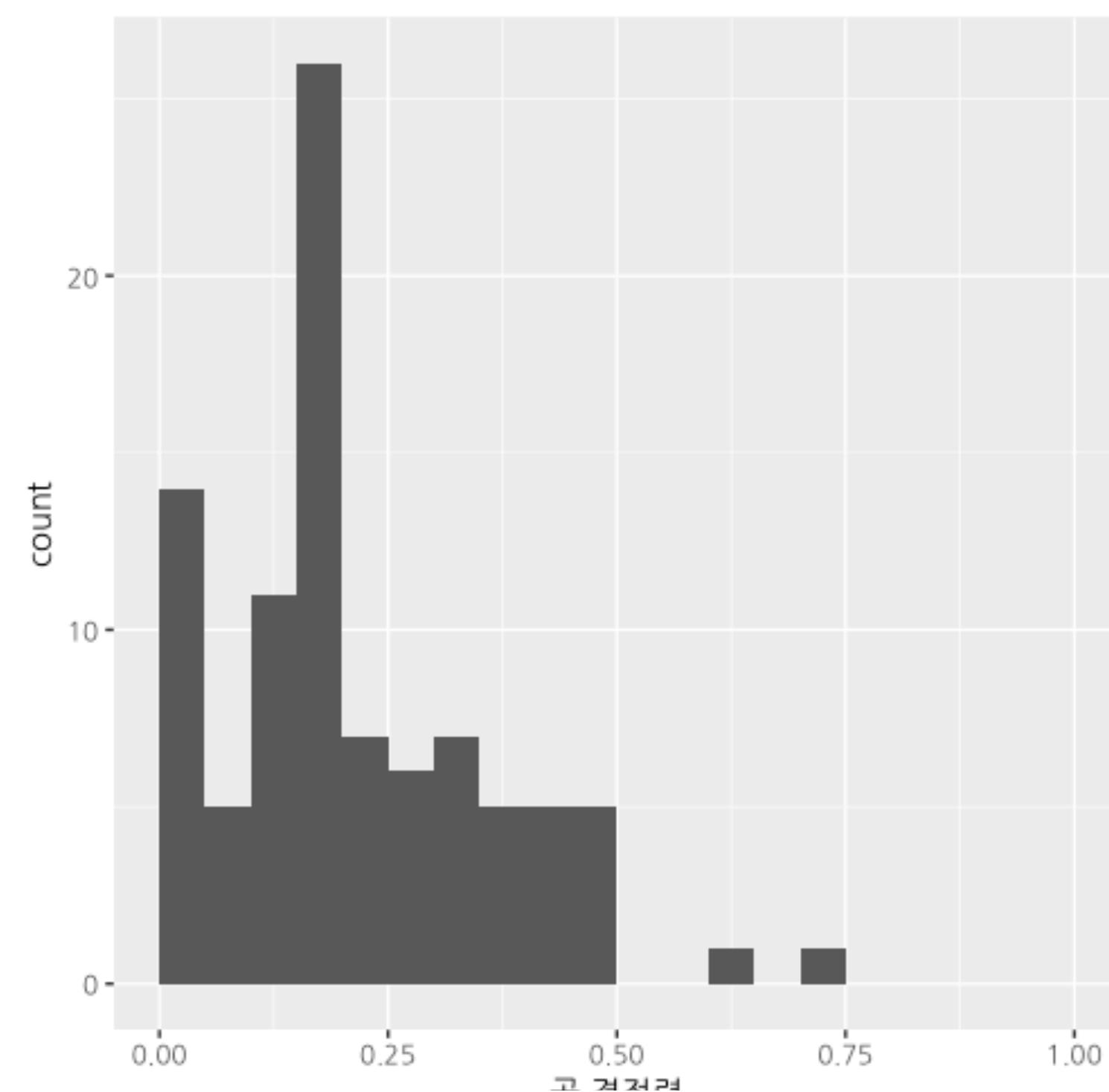
정보 하나.

슈팅 5회 이상의 선수들의 골 결정력 정보(분포) 를 사용, 평균 21.5%.

[대한민국 국대 전체 선수의 골 결정력]



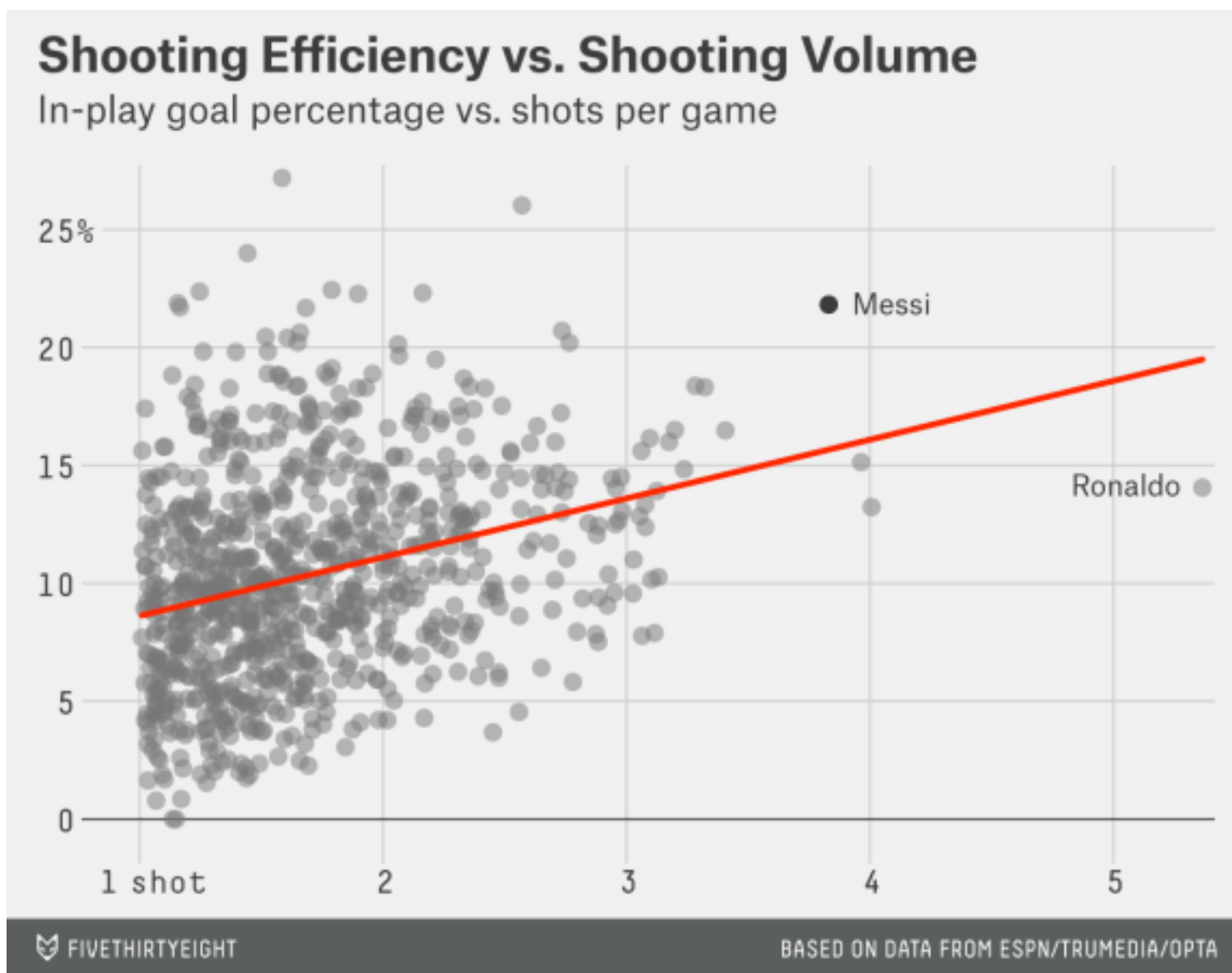
[대한민국 국대 슈팅=5 선수의 골 결정력]



4.

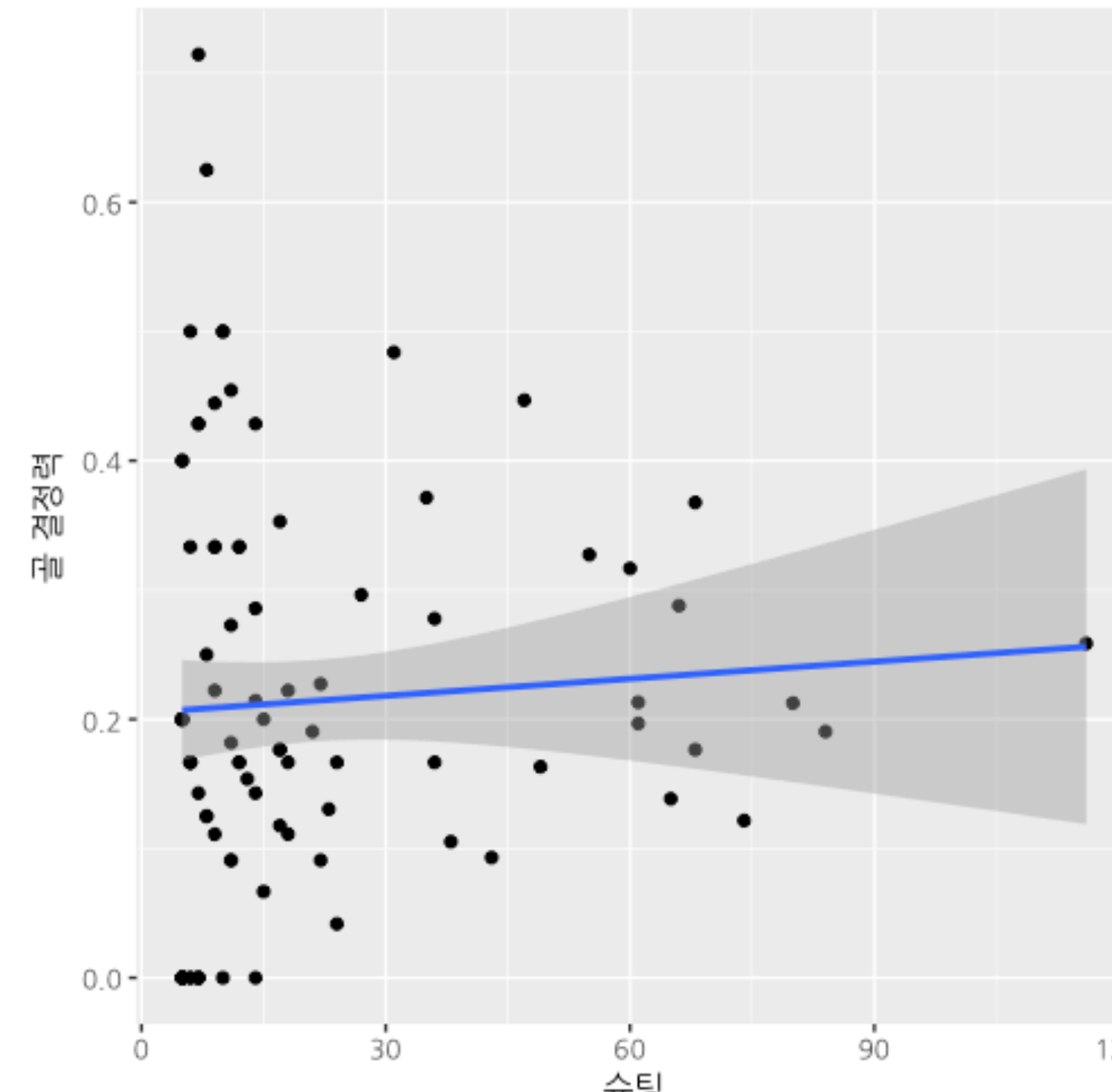
정보 둘.

1) Opta) 24,904 게임의 16,574 명 선수 데이터에 따르면, 슈팅을 많이 하는 선수의 슈팅 대비 골 성공률이 양의 상관관계.
2)대한민국 축구 국가 대표) 약하지만, 슈팅을 할 수록 슈팅 대비 골 성공률이 다소 향상



출처: FiveThirtyEight
(2014.7.1. "Lionel Messi Is Impossible")

[슈팅=5 슈팅수 x 골 결정력 분포]



5.

Results

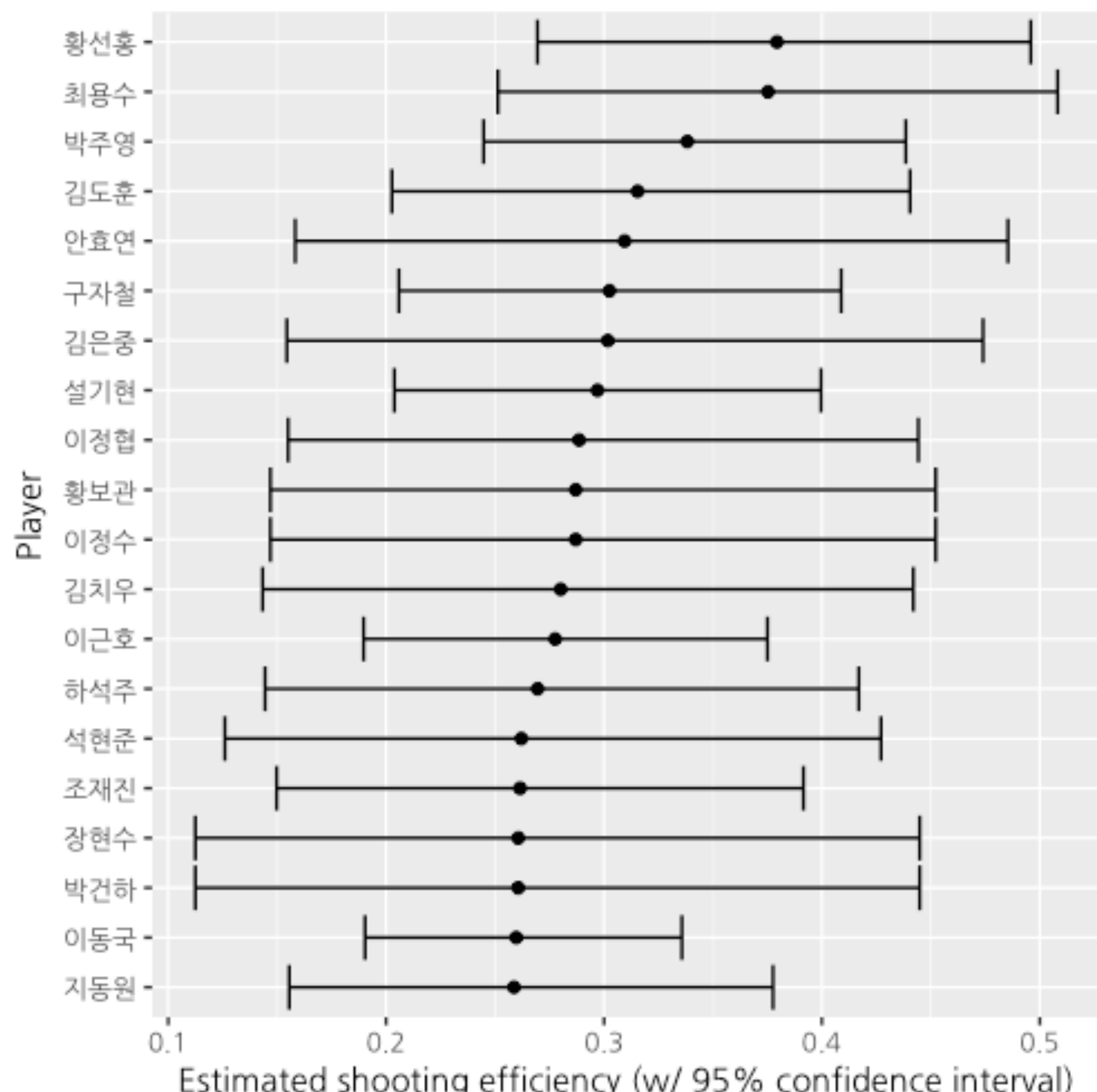
최고의 골 결정력 선수 Top 20

1)황선홍, 최용수, 박주영이 최고의 골 결정력을 보여준 선수, 이정수 선수는 수비수였음에도 Top10

2)골/슛은 황선홍 선수가 최용수 선수보다 더 낮지만 더 적극적으로 슈트를 하고 득점을 많았던 황선홍이 1위/ 슈트는 더 많이 했지만 상대적으로 성공률이 떨어지는 박주영이 3위

[사전 정보를 통해 다시 계산한 골 결정력 선수 Top20]

이름	골	슛	출장 경기수	골/슛	.fitted	.raw	.low
황선홍	21	47	39	44.68%	37.93%	26.95%	49.58%
최용수	15	31	28	48.39%	37.53%	25.13%	50.82%
박주영	25	68	60	36.76%	33.82%	24.48%	43.85%
김도훈	13	35	33	37.14%	31.54%	20.26%	44.05%
안효연	5	7	7	71.43%	30.94%	15.83%	48.54%
구자철	18	55	46	32.73%	30.25%	20.59%	40.89%
김은중	5	8	11	62.50%	30.17%	15.44%	47.40%
설기현	19	60	67	31.67%	29.70%	20.38%	39.96%
이정협	6	14	16	42.86%	28.86%	15.51%	44.43%
이정수	5	10	26	50.00%	28.70%	14.69%	45.22%
황보관	5	10	9	50.00%	28.70%	14.69%	45.22%
김치우	5	11	23	45.45%	28.01%	14.33%	44.20%
이근호	19	66	73	28.79%	27.76%	18.97%	37.50%
하석주	6	17	37	35.29%	26.95%	14.45%	41.69%
석현준	4	9	9	44.44%	26.21%	12.60%	42.71%
조재진	8	27	27	29.63%	26.15%	14.98%	39.15%
박건하	3	3	6	100.00%	26.07%	11.24%	44.49%
장현수	3	3	17	100.00%	26.07%	11.24%	44.49%
이동국	30	116	77	25.86%	25.97%	19.03%	33.57%
지동원	10	36	34	27.78%	25.87%	15.55%	37.76%



7.

What's Next?

1) 포지션에 따른 골 결정력의 차이?
2) 시기 (80년대, 90년대...), 대회 (월드컵/월드컵 예선/아시안컵/평가전)에 따른 골 결정력의 차이?

8.

References

1. EPL Stats Centre, <https://www.premierleague.com/stats>
2. FiveThirtyEight (2014.7.1.). Lionel Messi Is Impossible
3. Robinson (2017). Introduction to Empirical Bayes - Example from Baseball Statistics
4. Robinson (2017). ebb: Empirical Bayes on binomial data. R package version 0.1.
5. Rigby R.A. and Stasinopoulos D.M. (2005). Generalized additive models for location, scale and shape, (with discussion), Appl. Statist., 54, part 3, pp 507–554.
6. Wickham (2009). ggplot2: Elegant Graphics for Data Analysis. Springer–Verlag New York,

9.