



요인분석을 활용한 새로운 타격지표 제안

| 다변량 해석 7조 |

2017110525 송민철
2017110504 신준식
2018110476 배정민
2019110464 정유정

1

데이터 분석 배경 및 목적



타자의 능력 평가

기존의 평가방식



홈런

>

Power



타율

>

Hitting

실제



홈런

>

Power

Hitting

타격 지표에는 타자의 여러 능력이 반영되기 때문에
타자의 능력에 대한 정확한 평가를 할 수 없음

1

데이터 분석 배경 및 목적



기존 타격 지표



홈런



홈에서 1루
도달시간



타율



Etc...

요인분석

새로운 선수 능력
평가 지표

Power

Contact

Speed

Etc...

1

데이터 분석 배경 및 목적

추출된 요인의 합리성 검증



기존지표



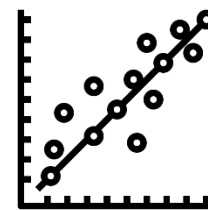
xwOBA
타자의 종합적인
타격능력 지표

뽑아낸 요인

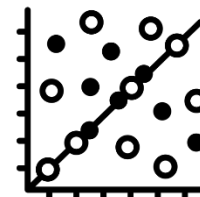


xwOBA
타자의 종합적인
타격능력 지표

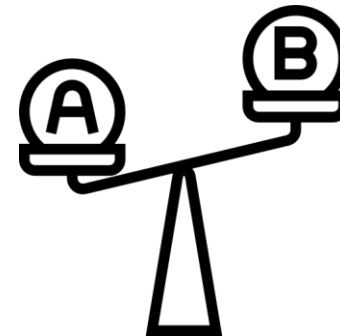
회귀분석



회귀모형 A



회귀모형 B



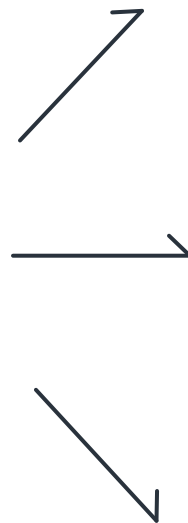
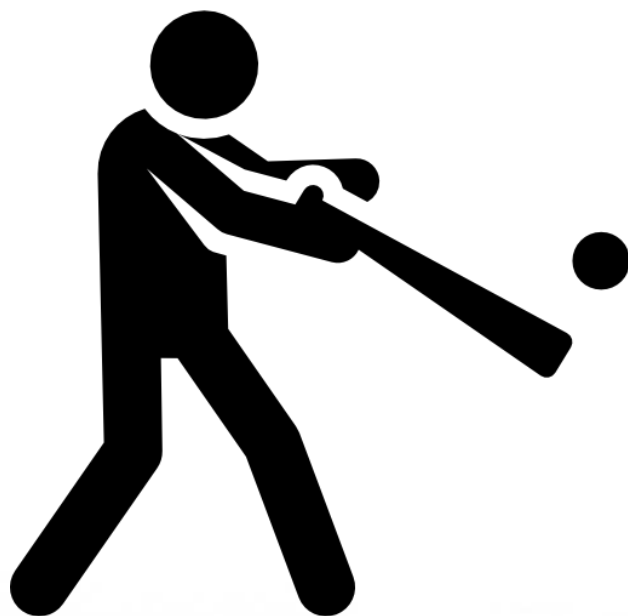
A와 B의 설명력(Adjusted R-squared) 비교



1

데이터 분석 배경 및 목적

요인들을 이용해서 선수에게 실제 적용



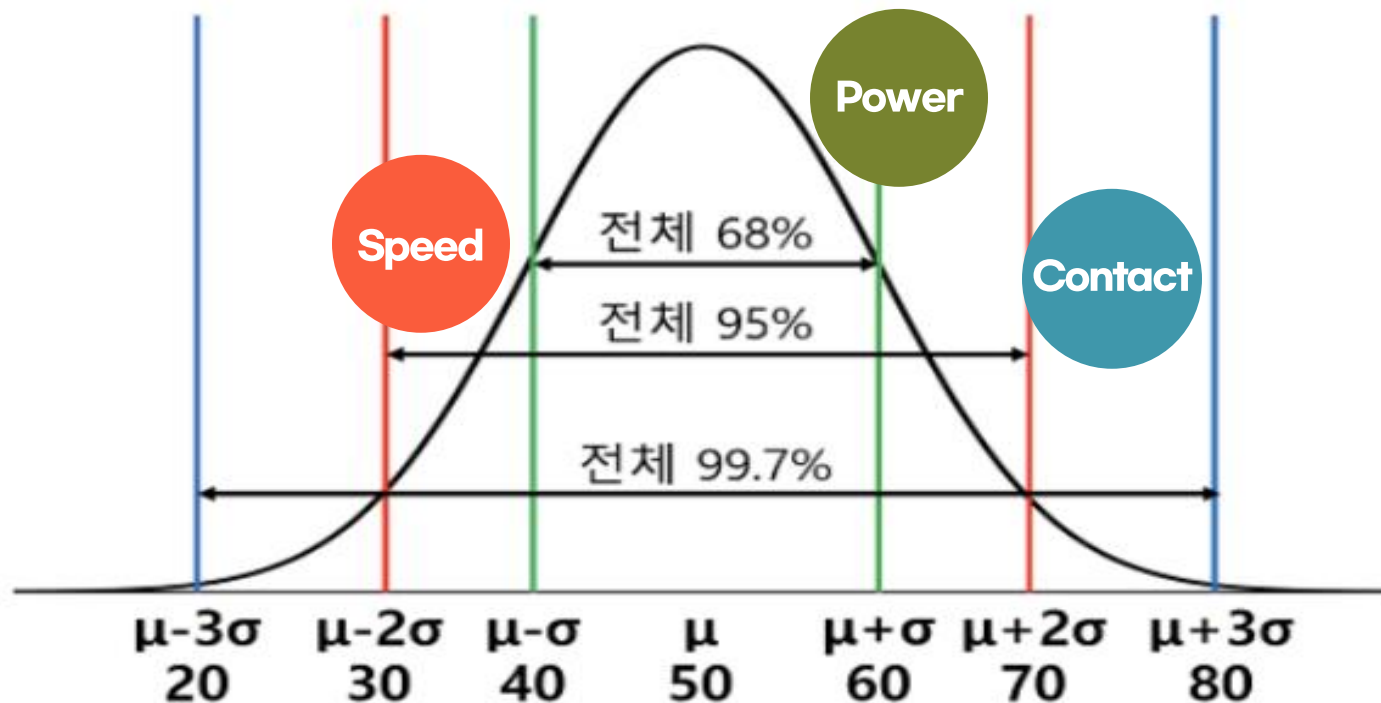
Power

Contact

Speed

1

데이터 분석 배경 및 목적



요인별 대표값의 Scale을 실제 야구 스카우팅 리포트에서 사용하는 측정 기준인 20-80 Scale로 변환



Stat0525

변수명	변수 설명 (단위)	변수 유형
last_name	last name	character
first_name	first name	character
player_id	선수별 구분을 위한 ID	numeric
year	연도	numeric
b_total_pa	타석 수	numeric
b_total_pitches	상대 투수 투구 수 총합	numeric
xba	기대 타율	numeric
xwoba	기대 가중 출루율	numeric
xobp	기대 출루율	numeric
xiso	기대 순수 장타율	numeric
exit_velocity_avg	출구 속도 (miles for hour)	numeric
launch_angle_avg	발사 각도 (°)	numeric
sweet_spot_percent	스윙 스팟에 맞출 확률 (%)	numeric
barrel_batted_rate	배럴 타구 확률 (%)	numeric
z_swing_percent	스트라이크 존 안에 들어오는 공에 스윙할 확률 (%)	numeric
oz_swing_percent	스트라이크 존 밖에 나가는 공에 스윙할 확률 (%)	numeric
oz_contact_percent	스트라이크 존 밖에 나가는 공을 컨택할 확률 (%)	numeric
iz_swing_percent	스트라이크 존 안에 들어오는 공을 컨택할 확률 (%)	numeric
n_bolts	최고 스프린트 스피드가 30ft/sec 이상이었던 횟수	numeric
hp_to_1b	홈에서 1루까지 걸리는 시간 (초)	numeric
sprint_speed	최고 스프린트 스피드 (feet for second)	numeric

※ 2015년 부터 2021년 까지(단축 시즌이었던 2020년 제외) 250타석 이상의 기록을 가진 타자의 데이터만 사용



FanGraphs_wrc

변수명	변수 설명 (단위)	변수 유형
Season	연도	numeric
Name	Full name	character
HR	홈런 개수	numeric
K%	삼진 확률 (%)	numeric
BABIP	인플레이 타율	numeric

※ 2015년 부터 2021년 까지(단축 시즌이었던 2020년 제외) 250타석 이상의 기록을 가진 타자의 데이터만 사용

FanGraphs_value

변수명	변수 설명 (단위)	변수 유형
Season	연도	numeric
Name	Full name	character
Batting	평균에 비해서 타격으로 얻은 점수	numeric
Base running	평균에 비해서 주루로 얻은 점수	numeric

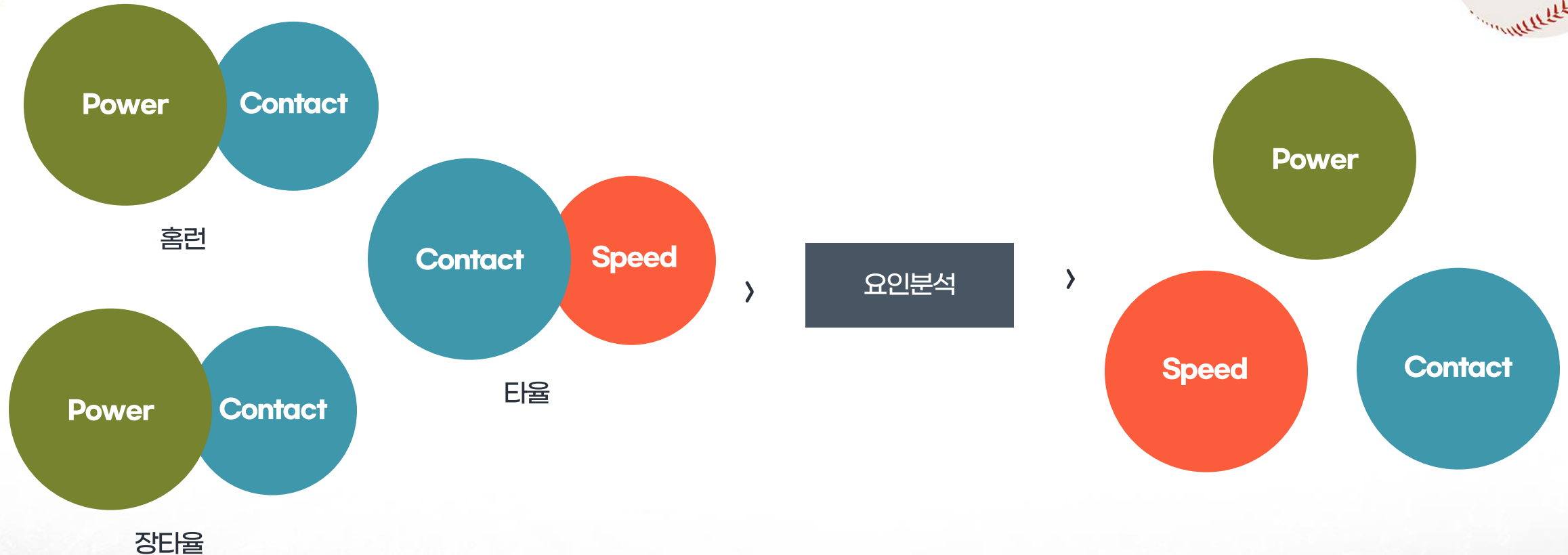
※ 2015년 부터 2021년 까지(단축 시즌이었던 2020년 제외) 250타석 이상의 기록을 가진 타자의 데이터만 사용



최종적으로 분석에 사용할 데이터 (1863행 23열)

xba	xobp_iso	xiso	exit_velocity_avg	launch_angle_avg	sweet_spot_percent	barrel_batted_rate	z_swing_percent	1 - oz_swing_percent	oz_contact_percent
0.259	0.070	0.212	89.3	11.2	0.327	0.104	0.641	0.745	0.488
0.214	0.110	0.169	90.4	10.2	0.305	0.097	0.679	0.773	0.371
0.263	0.071	0.153	88.8	11.9	0.340	0.031	0.682	0.776	0.599
0.231	0.065	0.126	89.1	10.7	0.285	0.046	0.707	0.764	0.576
0.248	0.114	0.181	86.7	11.7	0.300	0.075	0.616	0.797	0.593
0.262	0.120	0.204	89.7	13.2	0.342	0.088	0.651	0.817	0.650

iz_contact_percent	n_bolts	1.1*max - hp_to_1b	sprint_speed	HR	1-K	BABIP	ball_per_pa	Name	Season	xwoba	Batting	Base.Running
0.788	1	1.265	28.9	19	0.748	0.328	3.866505	AaronAltherr	2017	0.341	10.4	-1.9
0.775	1	1.285	29.0	8	0.681	0.247	4.052632	AaronAltherr	2018	0.317	-9.0	-1.2
0.848	5	1.455	28.4	11	0.831	0.285	3.871795	AaronHicks	2015	0.329	-1.9	0.4
0.807	5	1.495	28.5	8	0.812	0.248	3.825485	AaronHicks	2016	0.288	-15.7	-0.4
0.779	3	1.505	28.5	15	0.814	0.290	4.108033	AaronHicks	2017	0.348	12.3	2.6
0.795	7	1.505	28.1	27	0.809	0.264	4.280551	AaronHicks	2018	0.372	19.8	7.0



각 지표에 여러가지 **요인**이 들어가고
지표의 종류가 많아서 **다중공선성**이 높음

다중공선성 문제를 해결하고
선수 평가를 위한 핵심 **요인**만 추출



Uniquenesses

Uniquenesses:

xba	xobp_iso	xiso	exit_velocity_avg	launch_angle_avg
0.036	0.204	0.011	0.298	0.326
sweet_spot_percent	barrel_batted_rate	z_swing_percent	1 - oz_swing_percent	oz_contact_percent
0.135	0.058	0.074	0.219	0.270
iz_contact_percent	n_bolts	1.1*max - hp_to_1b	sprint_speed	HR
0.191	0.726	0.141	0.045	0.289
1-K	BABIP	ball_per_pa		
0.005	0.429	0.272		

7개의 **요인**이 Uniquenesses < 0.3 인 변수들의 **변동성**을 잘 설명하고 있다고 판단



launch_angle_avg, n_bolts, BABIP 변수를 제외하고 분석 진행

0.326

0.726

0.429





요인분석 결과

Loadings:

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
xba	0.410	0.619			0.544	0.205	
xobp_iso				0.935			0.145
xiso	0.965						
exit_velocity_avg	0.774	-0.129	-0.140		0.273	-0.192	
launch_angle_avg	0.267				-0.643	0.389	
sweet_spot_percent					0.130	0.946	
barrel_batted_rate	0.874	-0.266					
z_swing_percent							0.910
1 - oz_swing_percent	-0.114	0.160		0.915			
oz_contact_percent	-0.101	0.793					-0.106
iz_contact_percent	-0.131	0.777					-0.233
nBolts			0.478				
1.1*max - hp_to_1b			0.920				
sprint_speed			0.959				
HR	0.919	0.168	0.107		-0.166		
1-K		1.024		0.133			0.120
BABIP	-0.110	-0.188	0.131		0.696	0.274	
ball_per_pa		-0.307		0.488		0.134	-0.308

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
SS loadings	3.459	2.958	2.060	1.986	1.340	1.249	1.037
Proportion Var	0.192	0.164	0.114	0.110	0.074	0.069	0.058
Cumulative Var	0.192	0.356	0.471	0.581	0.656	0.725	0.783



요인분석 결과 해석

요인	밀접하게 관련된 변수	해석
Factor1	xiso, HR, barrel_batted_rate, exit_velocity_avg	Power
Factor2	1-K, oz_contact_percent, iz_contact_percent, xba	Contact
Factor3	1.1*max - hp_to_lb, sprint_speed	Speed
Factor4	xobp_iso, 1 - oz_swing_percent	Patient + Eye
Factor5	xba	Batting Skill
Factor6	sweet_spot_percent	Batting Skill For Extra Base Hit
Factor7	z_swing_percent	Aggressive + Eye



기존 지표들을 설명변수로 하는 회귀모형

Call:

```
lm(formula = logoba ~ xba + `1.1*max - hp_to_1b` + HR, data = aa0530)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.184449	-0.039375	-0.001107	0.038618	0.245018

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-1.7661820	0.0149697	-117.984	<2e-16	***
xba	2.4159417	0.0549664	43.953	<2e-16	***
`1.1*max - hp_to_1b`	-0.0566985	0.0062891	-9.015	<2e-16	***
HR	0.0056006	0.0001414	39.597	<2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05745 on 1859 degrees of freedom

Multiple R-squared: 0.7351, Adjusted R-squared: 0.7347

F-statistic: 1720 on 3 and 1859 DF, p-value: < 2.2e-16

Adjusted R-squared = 0.7347



뽑아낸 요인들을 설명변수로 하는 회귀모형

Call:
lm(formula = xwoba ~ power + `patient+eye` + contact + `batting skill` +
`aggressive+eye` + `batting skill for extra base hit`, data = matrix0520_23)

Residuals:
Min 1Q Median 3Q Max
-4.6394 -0.9957 -0.0505 0.9151 7.3468

Coefficients:

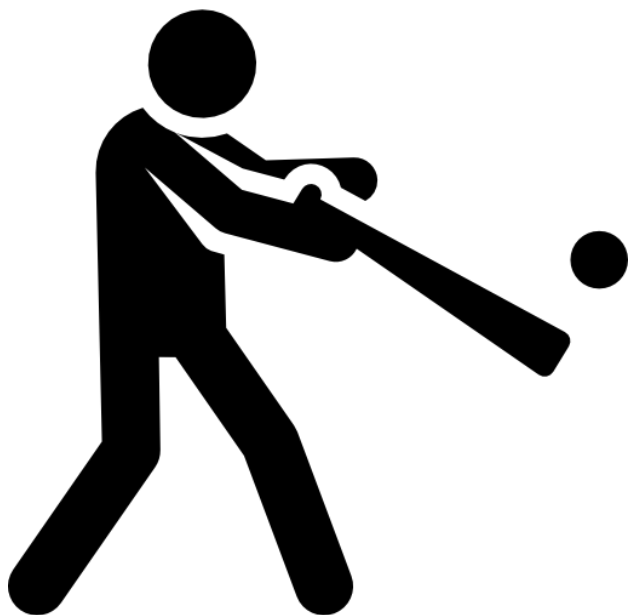
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-48.556983	0.493168	-98.46	<2e-16 ***
power	0.773705	0.004111	188.21	<2e-16 ***
`patient+eye`	0.338482	0.004345	77.90	<2e-16 ***
contact	0.358579	0.003899	91.97	<2e-16 ***
`batting skill`	0.292659	0.003537	82.74	<2e-16 ***
`aggressive+eye`	0.072731	0.004093	17.77	<2e-16 ***
`batting skill for extra base hit`	0.134983	0.003773	35.78	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.479 on 1856 degrees of freedom
Multiple R-squared: 0.9782, Adjusted R-squared: 0.9781
F-statistic: 1.387e+04 on 6 and 1856 DF, p-value: < 2.2e-16

Adjusted R-squared = 0.9781

› 기존 지표들을 설명변수로 하는 회귀모형의 설명력 0.7347 보다 높으므로 합리적이다



Power	$\{(x_{iso}-0.17) \times 13.70 + (\text{exit velocity avg} - 88.56) \times 0.01 + (\text{barrel batted rated} - 0.07) \times 3.67 + (HR - 16.19) \times 0\} \times 10 + 50$
Contact	$\{(x_{ba} - 0.25) \times 2.55 + (\text{oz contact percent} - 0.59) \times 0.36 + (\text{iz contact percent} - 0.83) \times 0.94 + ((1 - k) - 0.79) \times 14.52\} \times 10 + 50$
Speed	$\{(n_bolts - 4.96) \times 0.00 + \{(1.1 \times 5.15[- \text{hptolb}] - 1.23) \times 1.14\} + (\text{sprint speed} - 27.03) \times 0.52\} \times 10 + 50$
Patient + Eye	$\{(XOBP \text{ ISO} - 0.07) \times 19.19 + \{(1 - \text{oz swing percent}) - 0.72\} \times 7.08\} \times 10 + 50$
Batting Skill	$\{(x_{ba} - 0.25) \times 45.55\} \times 10 + 50$
Batting Skill for Extra Base Hit	$\{(\text{launch angle avg} - 12.43) \times 0.05 + (\text{sweet spot percent} - 0.33) \times 20.06\} \times 10 + 50$
Aggressive + Eye	$\{(z \text{ swing percent} - 0.67) \times 16.31\} \times 10 + 50$

20-80 Scale을 따르는 타자의 능력 평가 지표 생성

› 직관적으로 이해하기 쉽고 타자의 장단점과 특징을 파악하기에 유용





타자의 Power
홈런 개수? 장타율? 발사각?
배럴타구 비율?

추상적인 타자의 능력

타자의 Power

$$\{(x_{iso}-0.17) \times 13.70 + (exit\ velocity\ avg - 88.56) \times 0.01 \\ + (barrel\ batted\ rated - 0.07) \times 3.67 + (HR - 16.19) \times 0\} \times 10 + 50$$

구체적으로 수치화된 타자의 능력

	74	Power	51	
	44	Contact	51	
	59	Speed	56	
	34	Patient + Eye	55	
	61	Batting Skill	44	
	54	Batting Skill For Extra Base Hit	55	
	58	Aggressive + Eye	39	

오타니 쇼헤이(2022)

김하성(2022)

타자간 객관적인 능력과 더불어 성향까지도 수치적으로 확인하여 비교가 가능



Thank You :)
