

2020 빅콘테스트

# KBO 잔여시즌

## 팀별 승률, 타율 및 방어율 예측



수박팀

김동우(zxcvbnm1997@hanmail.net)

김시현(tlgus626@hanmail.net)

김재홍(kimjae1119@naver.com)

박준근(pjg2755@naver.com)

황재원(dch222@naver.com)

# INDEX

- 01    타율 & 방어율  
      분석 방향
- 02    타율 예측
- 03    방어율 예측
- 04    승률 예측

# 타율/방어율 분석 방향

대회 문제  
타율/ 방어율 예측

예측을 위한  
야구 데이터의 특징 파악

데이터의 특성에 맞는  
분석 방향 설정

## 1.1. 야구 데이터의 특징: High irreducible error

[예시 기록]


	AB	HIT	AVG
주전급 선수 A	5	10	0.5
신인 선수 B	1	1	1

KIA 브룩스 "가족 살아남은 게 기적...올해 복귀는 어려울 듯"

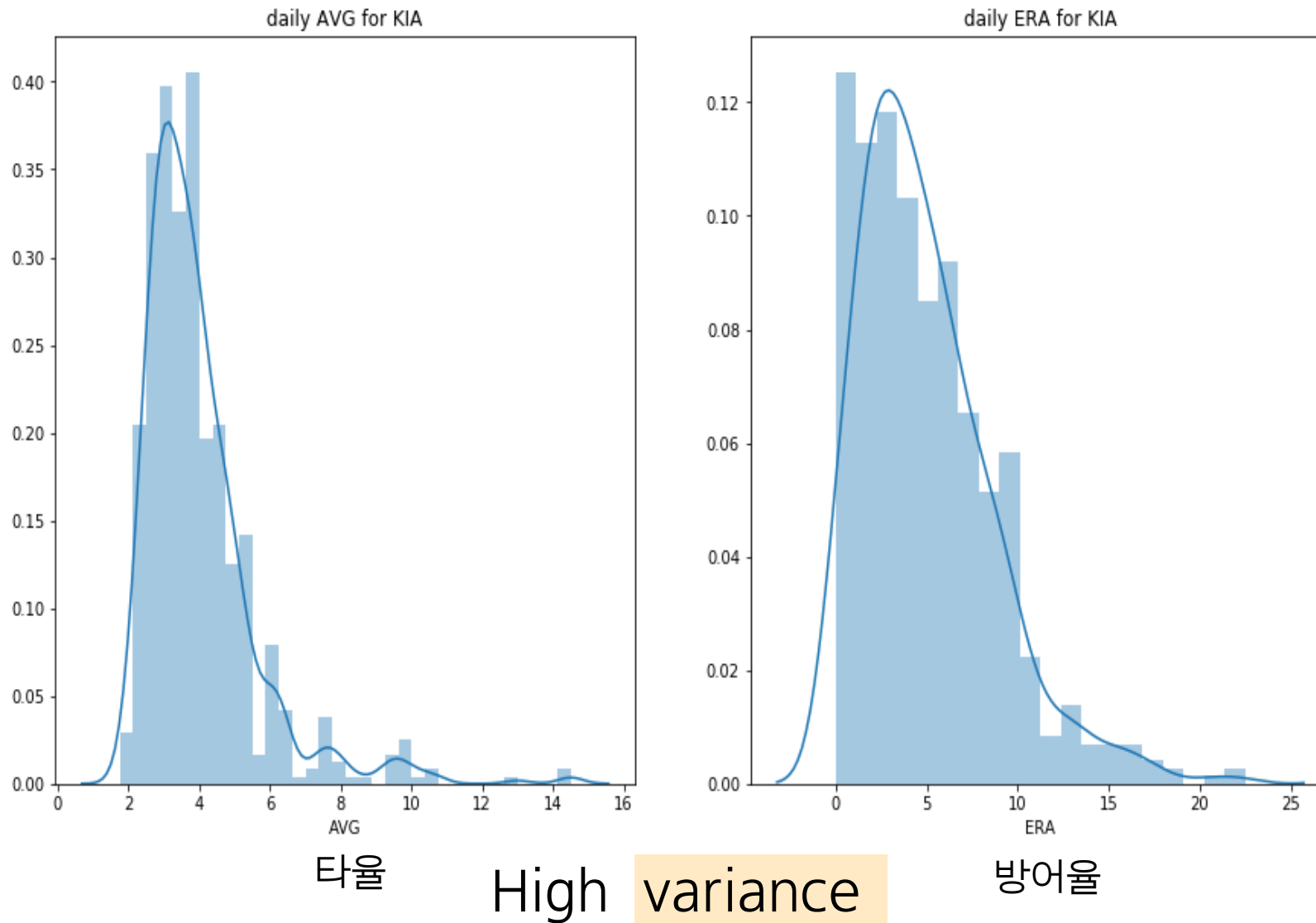
베테랑 10명 말소한 한화, 영건 대거 등록...7명은 시즌 첫 호출

선수 별 일별 기록으로는  
타율 예측 힘들

경기에 영향을 주는 요소 多

 Irreducible error가 크다고 할 수 있음

## 1.2. 야구 데이터의 특징: High variance

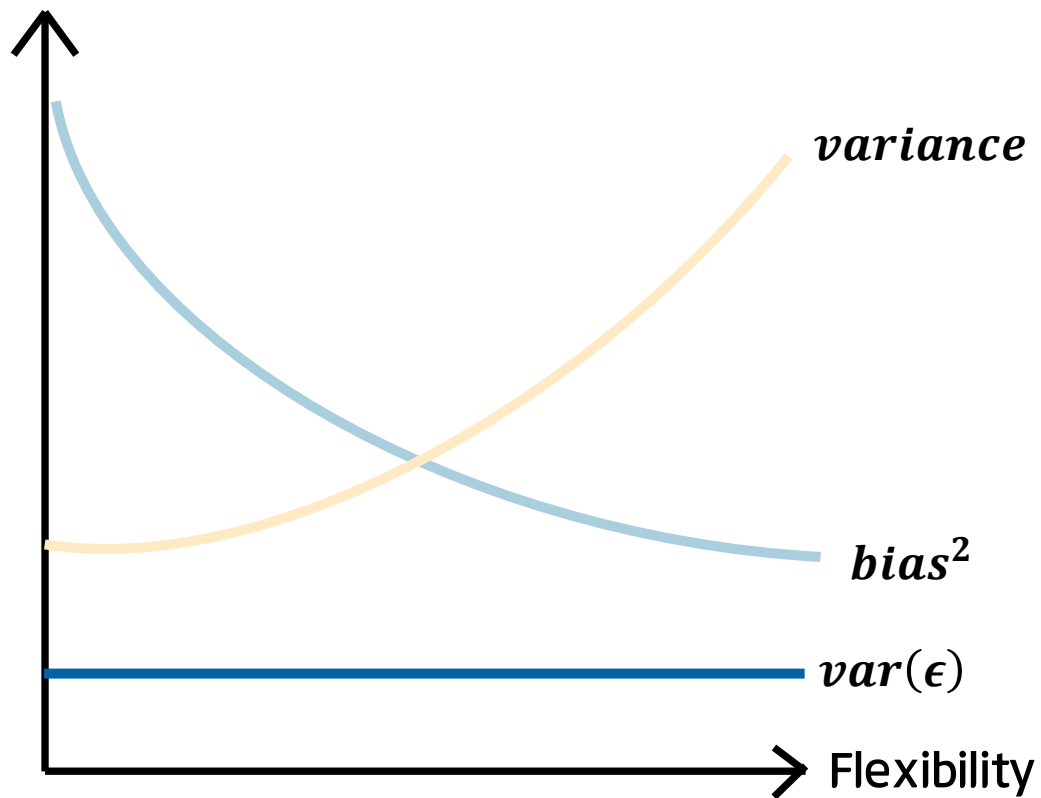


## 1.3. 데이터 특성에 맞는 분석 방향

MSE decomposition

$$E(Y - \hat{Y})^2 = \text{bias}^2 + \text{variance} + \text{var}(\epsilon)$$

Irreducible error



irreducible error와 variance가 크기 때문에  
Flexible model을 적용했을 때  
Overfitting 가능성 존재



일부 bias를 감안하더라도  
Linear regression model 선정

## 1.3. 데이터 특성에 맞는 분석 방향

분산 안정화 방안

**WLS**

Weighted Linear Squares

$$sample\ weight = \frac{1}{var(residual)^2}$$

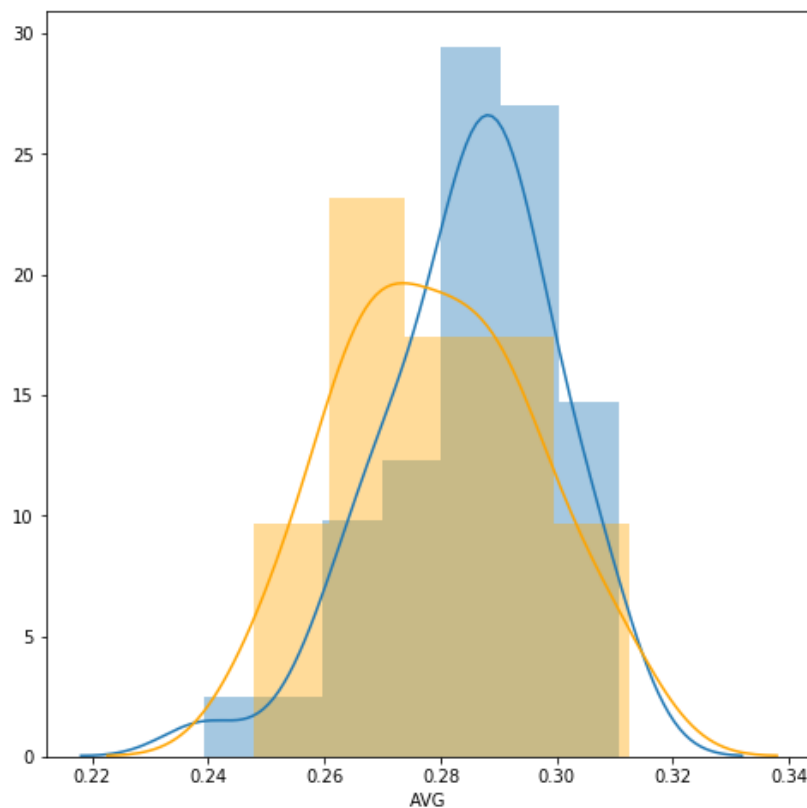
잔차가 큰 것에 적은 가중치, 작은 것에 큰 가중치 부여



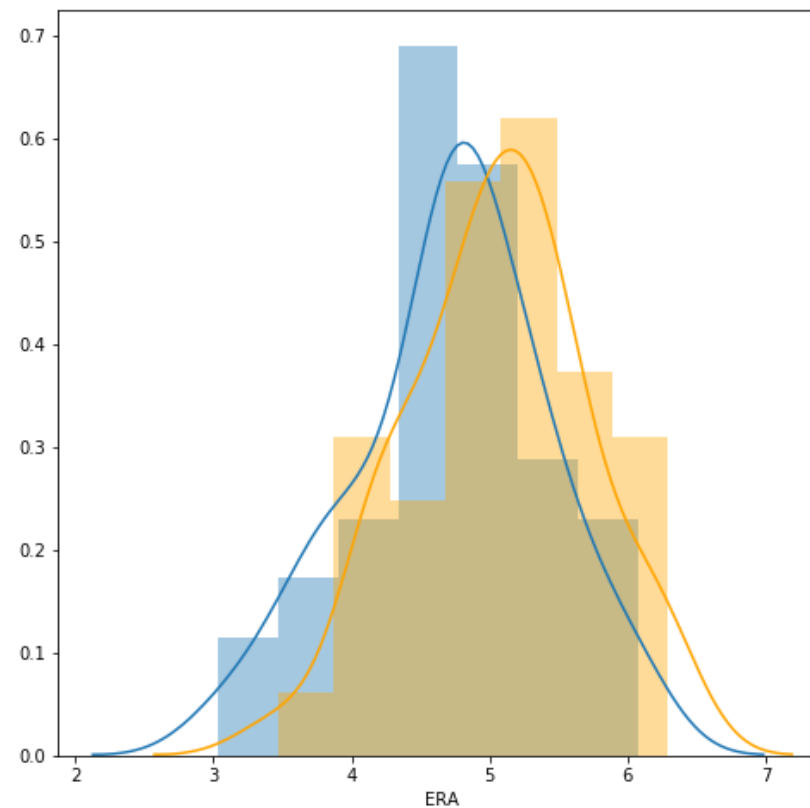
이분산성의 해결을 위해 쓰이는 방법

“Overfitting 가능성을 줄이고 분산을 안정화 할 수 있는 **WLS를 모델**로 선정”

## 1.3. 야구 데이터의 특징: 홈/ 원정 차이



타율



방어율

홈과 원정에서의 데이터 경향이 달랐기 때문에 나누어 분석 진행

☑ Column 'TB\_SC' : 경기의 초/말 구분 → 원정 경기는 초, 홈경기는 말에 진행하기 때문에 TB\_SC 기준으로 홈/원정 구분



## Chapter 2

# 타올

- 1) EDA & 데이터 전처리
- 2) Feature engineering
- 3) Data modeling

## Data set

팀 타자 데이터

## Feature engineering

## Modeling

## Predict

HOME  
TB\_SC="B"

X

각 시즌 1~115 경기간  
지표 누적 합  
+ 지표 추가

Y

115~144 경기  
팀 타율

- ① 변수의 유의성 T-test
- ② PCA

- ① WLS
- ② Outlier 제거

최종 예측  
BY 가중평균AWAY  
TB\_SC="T"

X

각 시즌 1~115 경기간  
지표 누적 합  
+ 지표 추가

Y

115~144 경기  
팀 타율

- ① 변수의 유의성 T-test
- ② PCA

- ① WLS
- ② Outlier 제거



2.1.1. 변수 추가

PA	HIT	AB	BB	IB	HP	SF	KK
타자	안타	총 타수	4구	고의 4구	사구	희비	삼진

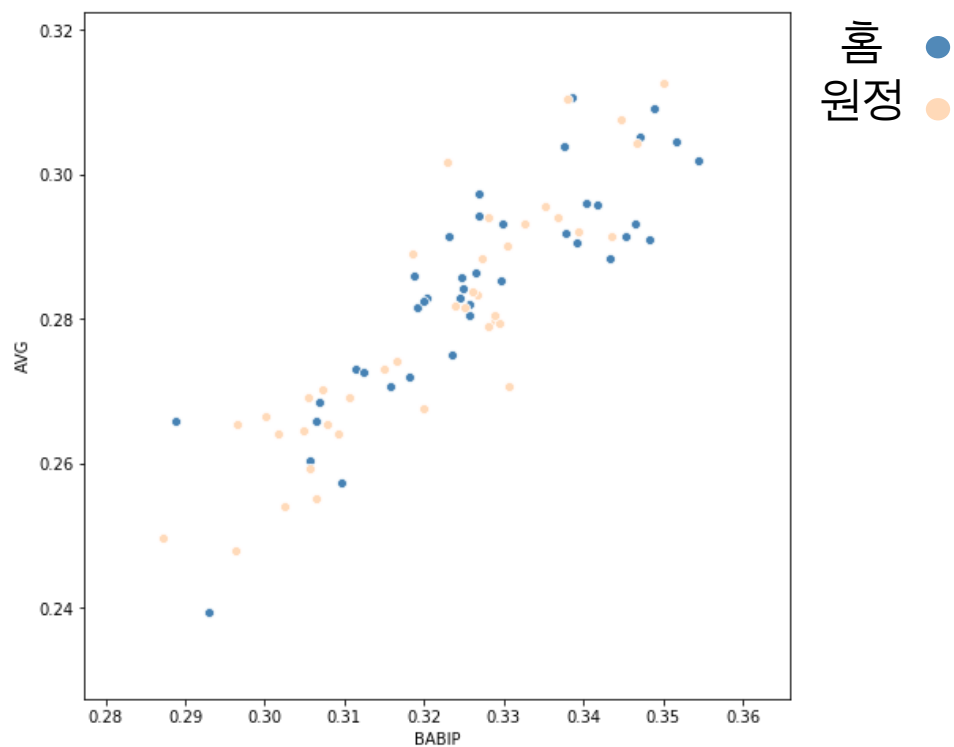


“기존 지표들을 활용하여 타율에 영향을 미치는 새로운 변수 추가”  
AVG

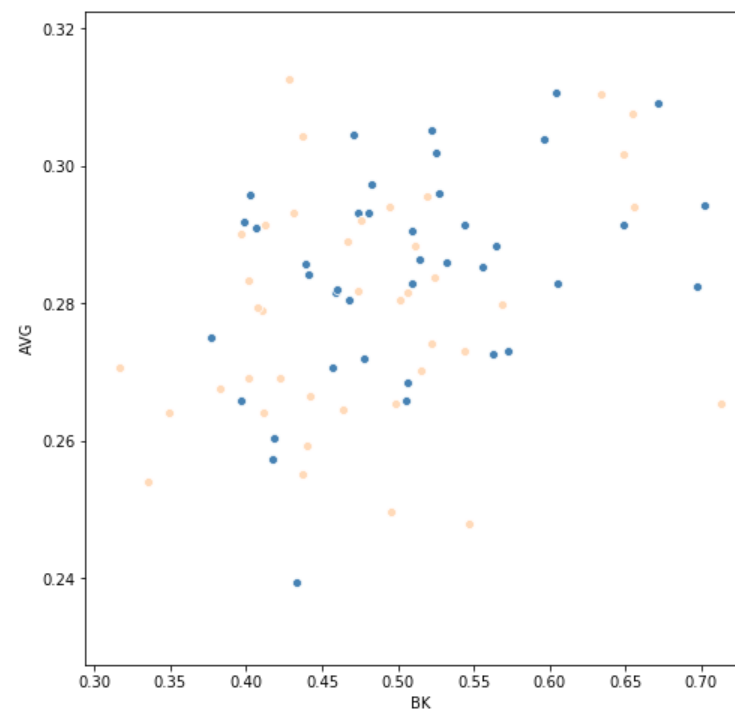
AVG	AB/HIT
OPS	$OPB+SLG=\{(HIT+BB+IB+HP)/(AB+BB+IB+HP+SF)\}+\{(H1+2*H2+3*H3+4*HR)/AB\}$
CQ	$\{0.5*HIT+0.3*(H1+2*H2+3*H3+4*HR)\}/(AB-KK)$
BABIP	$(HIT-HR)/(AB-KK-HR+SF)$
BK	BB/KK

## 2.1.1. 변수 추가

새로운 변수와 타율과의 관계



BABIP-AVG

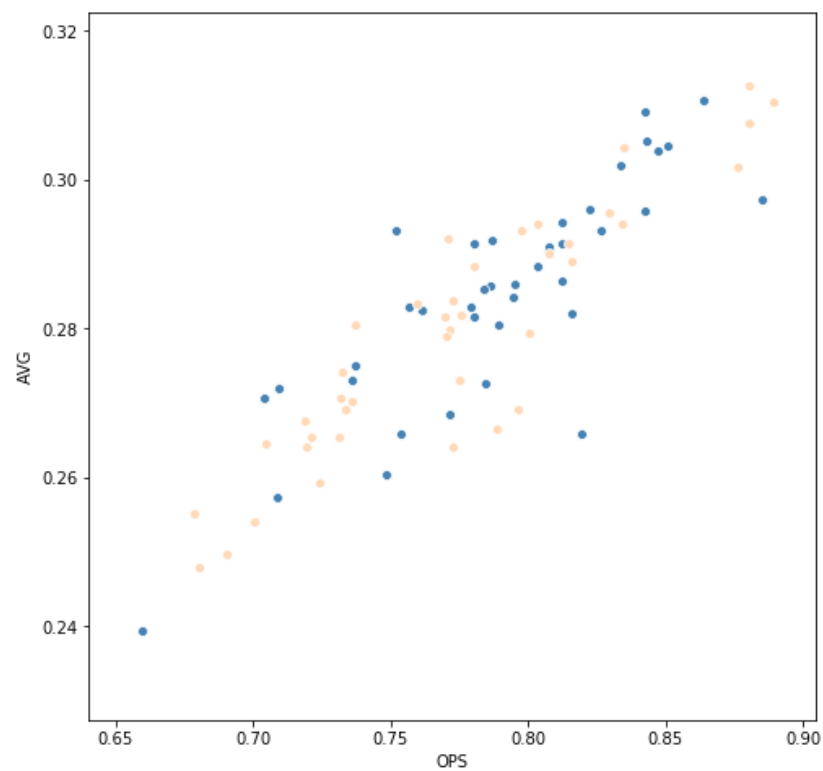


BK-AVG

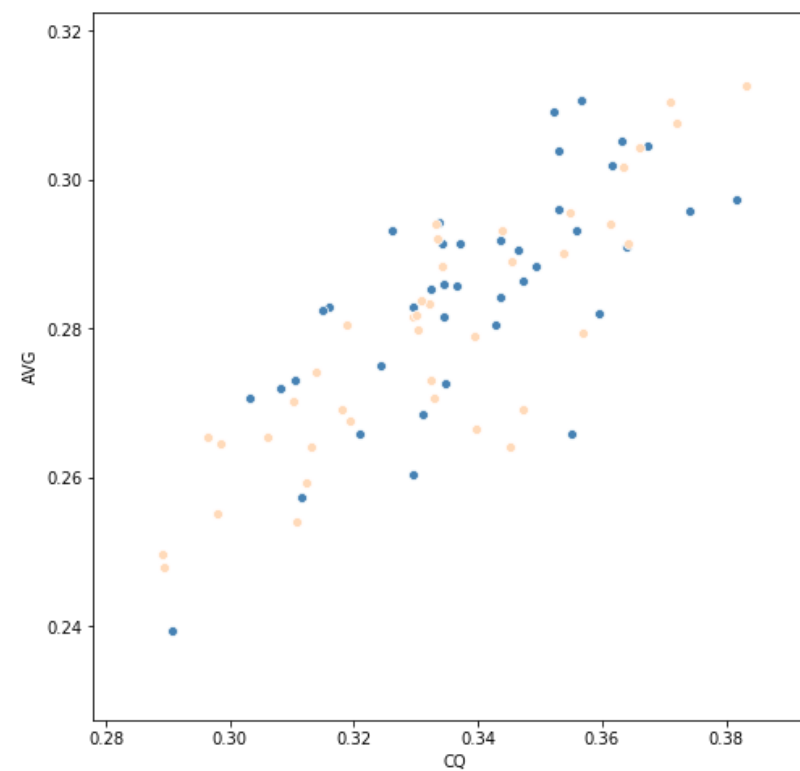
“타율 예측 변수로 사용 가능”

## 2.1.1. 변수 추가

새로운 변수와 타율과의 관계



OPS-AVG



CQ-AVG

“타율 예측 변수로 사용 가능”

## 2.1.2. 데이터 정리

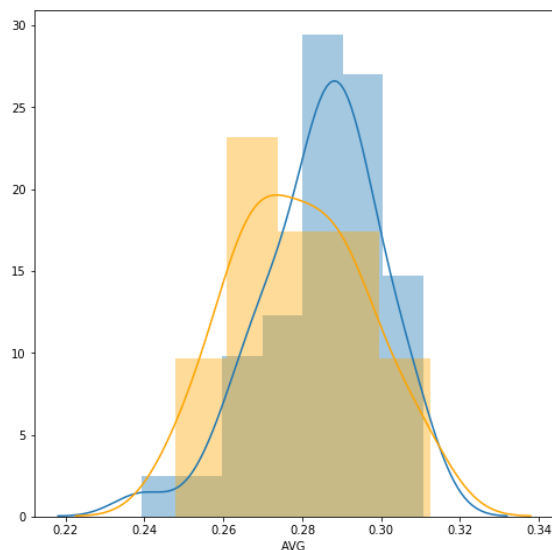
기록을 시즌별 팀 누적합으로 요약

home\_train\_x

	PA	AB	RBI	RUN	HIT	H2	H3	HR	SB	CS	...	GD	H1	AVG	OPS	CQ	BABIP	BK
<b>16HH</b>	2223	1946	305	323	576	105	11	58	29	12	...	50	402	0.295992	0.822161	0.353043	0.340342	0.527273
<b>16HT</b>	2160	1889	295	316	541	105	11	60	39	24	...	36	365	0.286395	0.812171	0.347156	0.326544	0.514589
<b>16KT</b>	2246	1973	257	274	539	89	4	45	40	19	...	56	401	0.273188	0.736132	0.310477	0.311475	0.572222
<b>16LG</b>	2227	1931	296	319	563	87	19	35	57	27	...	29	422	0.291559	0.780115	0.337054	0.345324	0.543590
<b>16LT</b>	2166	1865	274	294	538	104	5	49	76	29	...	38	380	0.288472	0.803302	0.349382	0.343399	0.564792
...											...							...
<b>19LT</b>	2130	1892	238	256	487	92	8	40	35	12	...	27	347	0.257400	0.708883	0.311565	0.309557	0.417062
<b>19NC</b>	2285	2014	288	308	576	102	11	61	34	18	...	50	402	0.285998	0.794945	0.334483	0.318688	0.531856
<b>19OB</b>	2241	1926	281	298	544	98	12	33	34	9	...	49	401	0.282451	0.761444	0.314815	0.319975	0.696697
<b>19SK</b>	2115	1865	263	272	496	82	6	56	43	17	...	31	352	0.265952	0.753643	0.320999	0.306407	0.505208
<b>19SS</b>	2220	1929	283	293	518	97	10	61	50	15	...	40	350	0.268533	0.771465	0.330971	0.306917	0.506173
<b>19WO</b>	2347	2067	316	335	590	114	25	42	46	9	...	46	409	0.285438	0.784043	0.332342	0.329723	0.555844

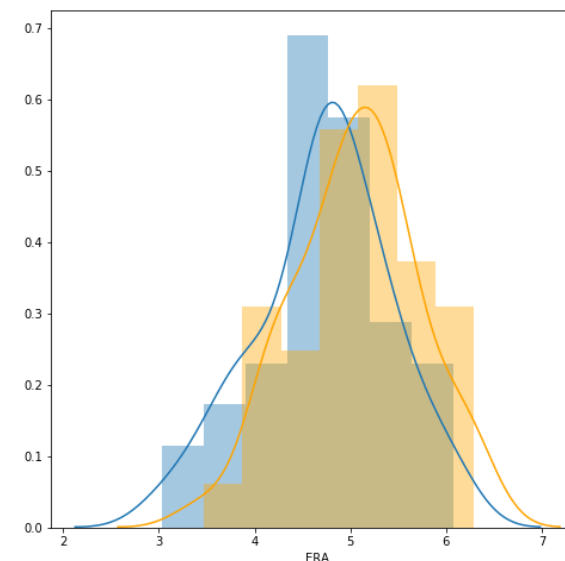
결과 TABLE

### 2.2.1. 홈-원정 간 편차 반영 변수 추가



홈  
원정

홈 대비  
원정 경기의  
분산은 큼



홈/원정에서의 실력 차이를 반영해야 하기 때문에 홈/원정간 편차를 반영해야 함

변수 추가

diff_PA	HOME PA-AWAY PA
diff_HIT	HOME HIT -AWAY HIT
diff_AB	HOME AB-AWAY AB

## 2.2.2. 변수별 T-test 시행

### “변수의 유의성 판단하기 위하여 T-test 시행”

독립 변수를 Scaling 해주었기 때문에 각 coefficient는 t test statistic임

HOME

Variable	Coefficient
CS	0.000149
IB	0.001735
RUN	0.002037
H3	0.003204
GD	0.005874
BB	0.008920

⋮

AWAY

Variable	Coefficient
IB	0.001535
diff_HIT	0.004765
H3	0.005478
GD	0.005514
CS	0.005520
H1	0.006235

⋮

Coefficient  $\leq 0.01$ 인  
변수 제거



## 2.2.3. PCA 시행

HOME		AWAY	
VIF	features	VIF	features
10.994	SH	593949.33	PA
4924.98	CQ	401569.82	AB
4438.75	AVG	3443.26	HIT
3171.67	PA	2729.51	CQ
2238.51	OPS	1980.83	BABIP
1278.95	KK	8.122	H2
⋮		⋮	

VIF 확인

High VIF

변수들 간 다중 공선성을 확인

“Feature Extraction: PCA를 통해 다중 공선성 해결”

## 2.3.1. DATA SET 설정

✓ 2016~2019 DATA

Train set

✓ 2020 DATA

Test set

HOME

Train set

Test set

AWAY

X: 각 연도별 1~115 경기

Y: 각 연도별 115~144 경기의 AVG

X: 2020년 1~115경기

Y: **2020년 잔여 경기의 AVG**

**: 예측해야 할 것**

팀별로 27~30의 잔여 경기  
: 마지막 30 잔여 경기를 예측하자!  
→ 115경기를 기준으로 나눔

## 2.3.1. DATA SET 설정

 **STATIZ**에서 크롤링!

홈/ 원정 팀별 타자 데이터  
Test set으로 활용



제공해주지 않은 2020 DATA는?



2020 DATA

Train set

Test set

HOME

Train set

Test set

AWAY

### 2.3.2. 홈/원정 데이터 분석 결과

#### 1) Sample weight 부여

$$\text{sample weight} = \frac{1}{\text{var}(\text{residual})^2}$$

#### 2) RMSE를 최소화하는 최적의 주성분 개수 (n\_component)

	n_component	RMSE
HOME	5	0.02717
AWAY	3	0.02474



HOME:0.917957

AWAY:0.817161

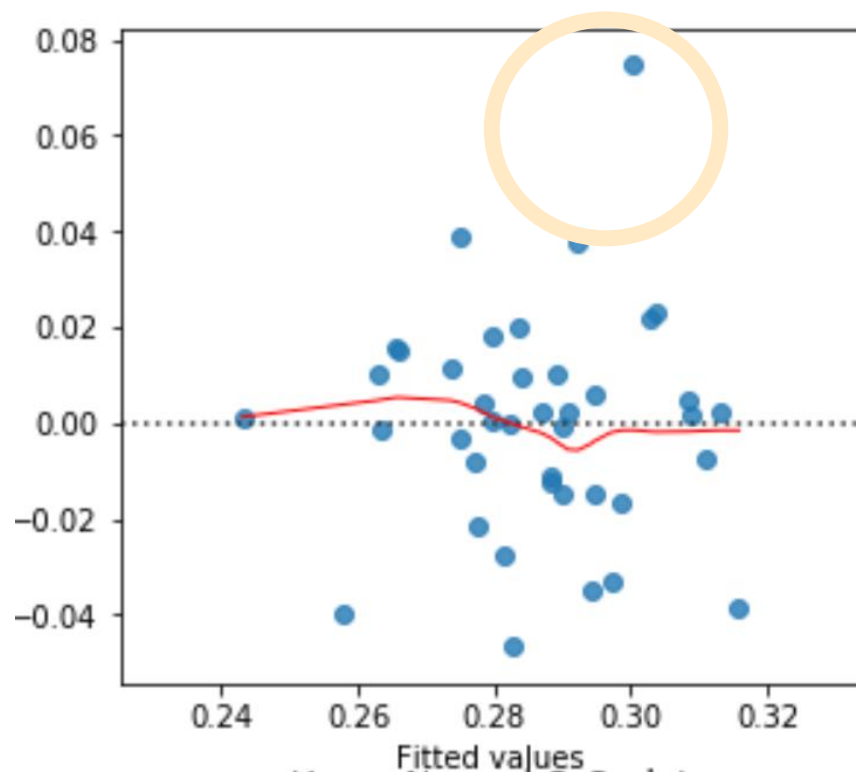
분산 설명력 확인

:홈/ 원정 모두 좋은 설명력 가짐

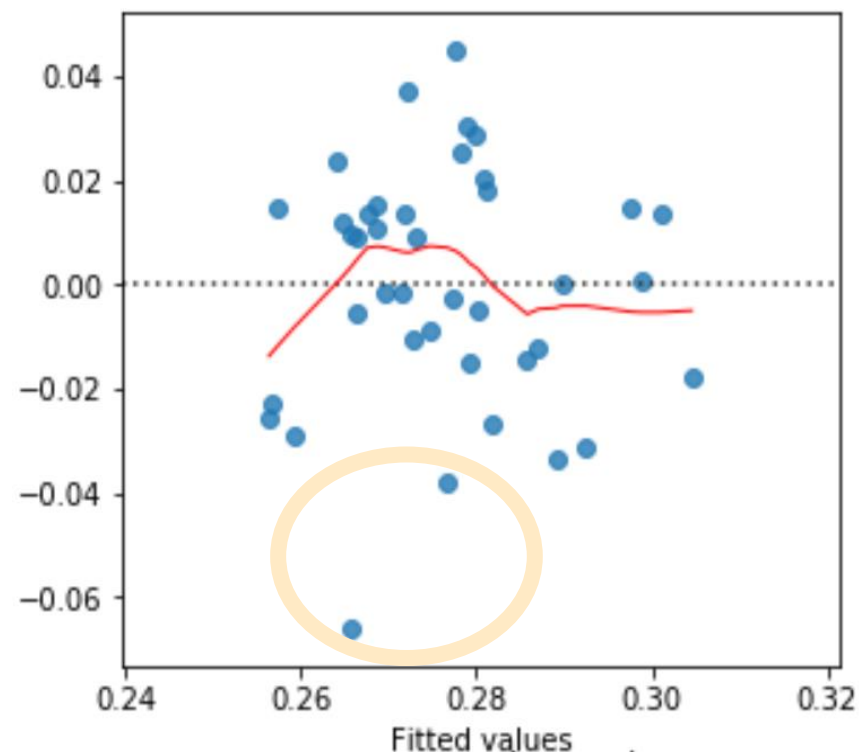
## 2.3.2. 홈/원정 데이터 분석 결과

### 3) Residual analysis : Outlier 제거

HOME



AWAY



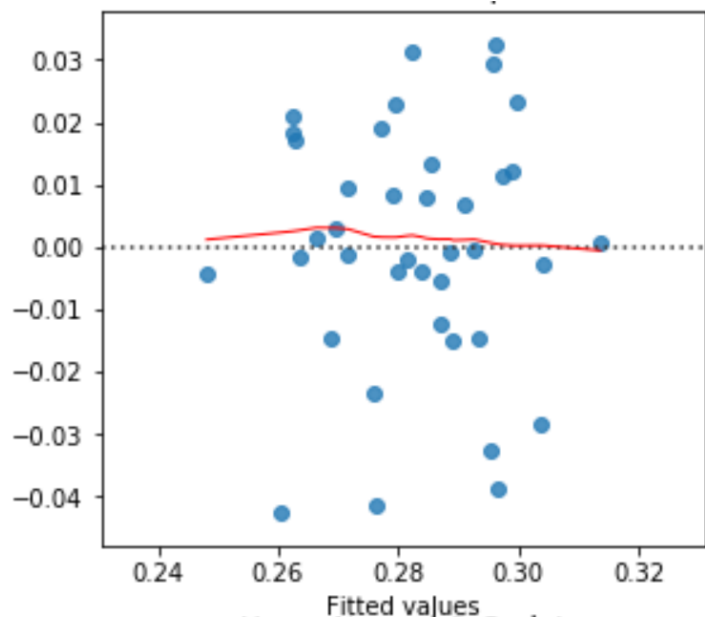
## 2.3.2. 홈/원정 데이터 분석 결과

### 3) Residual analysis : Outlier 제거

Outlier

“그래프에서 눈에 띄는 Outlier 제거”

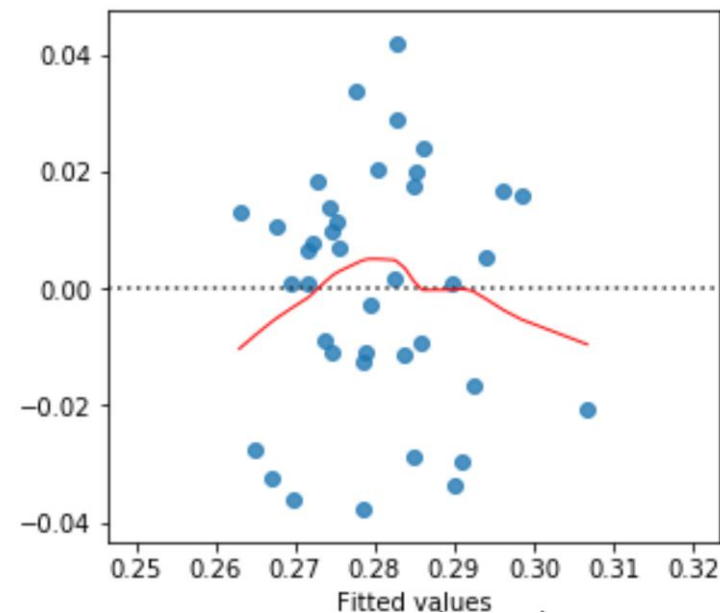
HOME



$R^2$   
0.999836  
→ 0.99876

RMSE  
0.023  
→ 0.0193

AWAY



$R^2$   
0.99999797  
→ 0.99999565

RMSE  
0.0227  
→ 0.0202

## 2.3.2. 홈/원정 데이터 분석 결과

### 3) Predict

잔여일 동안의 홈/ 원정 경기 수



홈/ 원정 예측값의 가중 평균

	T_ID	HOME_COUNT	AWAY_COUNT	HOME_PRED_AVG	AWAY_PRED_AVG	AVG
0	HH	14	12	0.239124	0.263918	0.251102
1	HT	15	14	0.280979	0.274756	0.278412
2	KT	13	14	0.291688	0.272311	0.281668
3	LG	14	11	0.268474	0.295624	0.282852
4	LT	14	14	0.270555	0.278311	0.274433
5	NC	15	13	0.301360	0.285105	0.295890
6	OB	11	14	0.260405	0.298721	0.288187
7	SK	14	10	0.257524	0.271381	0.269489
8	SS	10	16	0.267947	0.269154	0.283131
9	WO	9	11	0.263195	0.292749	0.283707

## 2.3.2. 홈/원정 데이터 분석 결과

### 3) Predict

최종 예측  
타울

	T_ID	AVG
5	NC	0.295890
6	OB	0.288187
9	WO	0.283707
8	SS	0.283131
3	LG	0.282852
2	KT	0.281668
1	HT	0.278412
4	LT	0.274433
7	SK	0.269489
0	HH	0.251102



## Chapter 3

# 방어울

- 1) EDA & 데이터 전처리
- 2) Feature engineering
- 3) Data modeling

## 방어율 분석 OUTLINE

Data set

팀 타자 데이터

Feature  
engineering

Modeling

Predict

HOME  
TB\_SC="B"

X

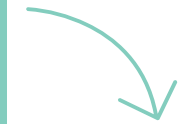
각 시즌 1~115 경기간  
지표 누적 합  
+ 지표 추가

Y

115~144 경기  
팀 자책점

- ① 변수의 유의성 T-test
- ② PCA

- ① WLS
- ② Outlier 제거

최종 예측  
BY 가중평균AWAY  
TB\_SC="T"

X

각 시즌 1~115 경기간  
지표 누적 합  
+ 지표 추가

Y

115~144 경기  
팀 자책점

- ① 변수의 유의성 T-test
- ② PCA

- ① WLS
- ② Outlier 제거



3.1.1. 변수 추가

✓ ERA=(ER\*3\*9)/INN2 → 팀 별 INN2수는 비슷하므로 ER 예측 후 최종 ERA 예측

ER	INN2	R	HP	BB	KK	AB	SF
자책점	이닝*3	실점	사구	4구	삼진	타수	희비

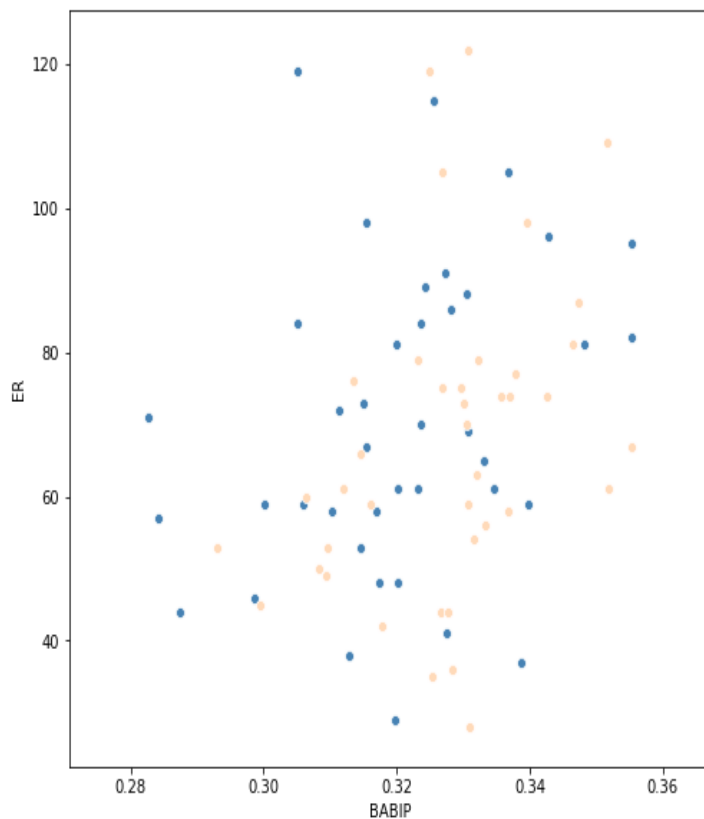


“기존 지표들을 활용하여 방어율에 영향을 미치는 새로운 변수 추가”

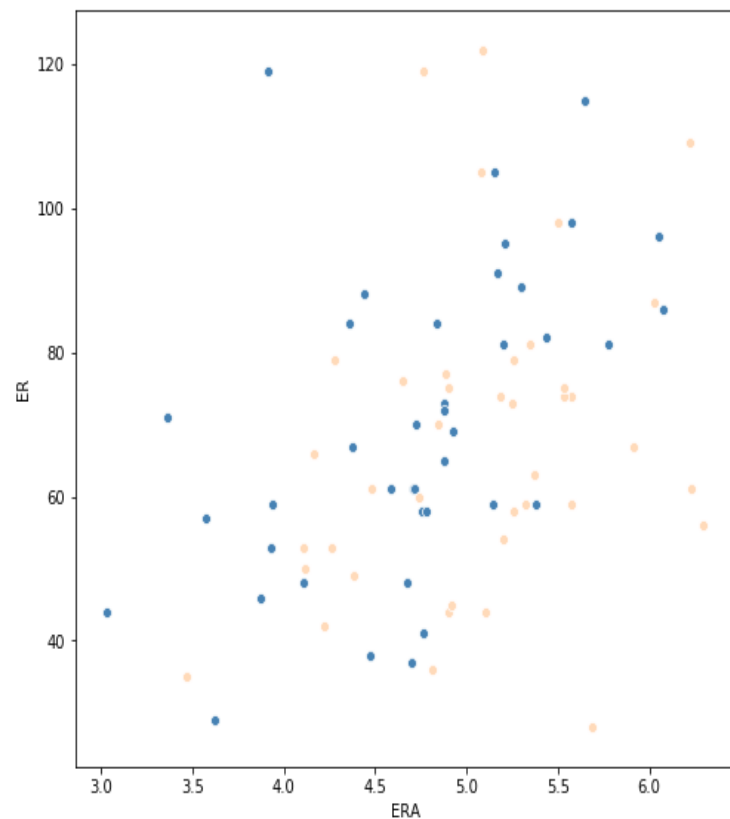
ERA	$(ER * 3 * 9) / INN2$
WHIP	$(BB + HIT) / (INN2 / 3)$
BABIP	$(BB + HIT) / (INN2 / 3)$
K/BB	$KK / BB$
LOB%	$(HIT + BB + HP - R) / (HIT + BB + HP - (1.4 * HR))$
OPS	$\{(HIT+BB+HP) / (AB+BB+HP+SF)\} + \{(H1+2*H2+3*H3+4*HR)/AB\}$
SCORE%	$ERA / (WHIP*9)$

## 3.1.1. 변수 추가

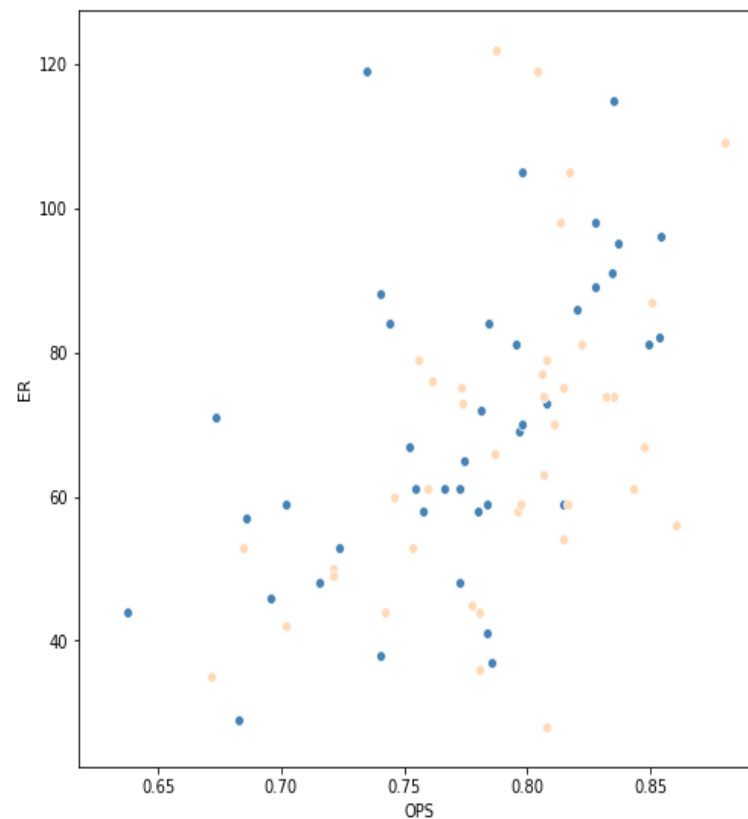
새로운 변수와 자책점과의 관계

홈 ●  
원정 ●

BABIP-ER



ERA-ER



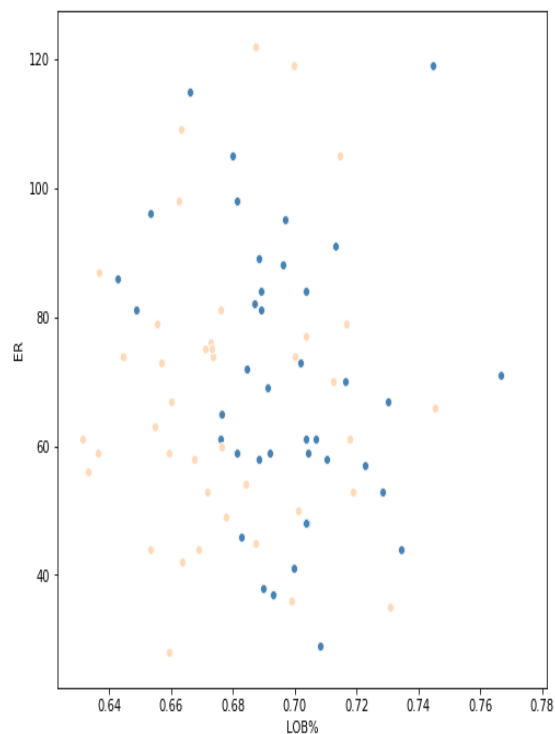
OPS-ER

“방어율 예측 변수로 사용 가능”

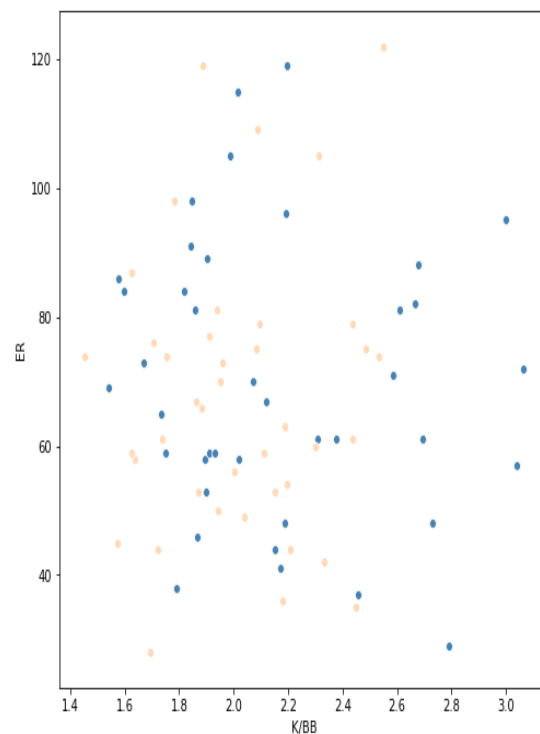
### 3.1.1. 변수 추가

새로운 변수와 자책점과의 관계

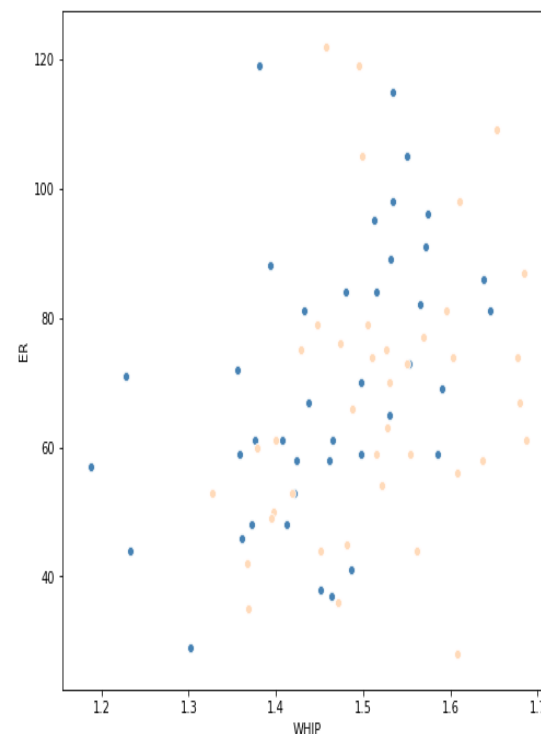
홈 ●  
원정 ●



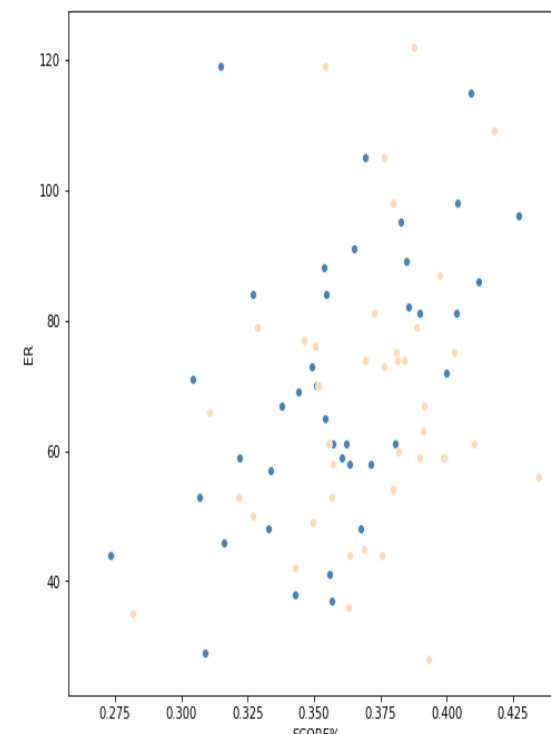
LOB%-ER



K/BB-ER



WHIP-ER



SCORE%-ER

“방어율 예측 변수로 사용 가능”

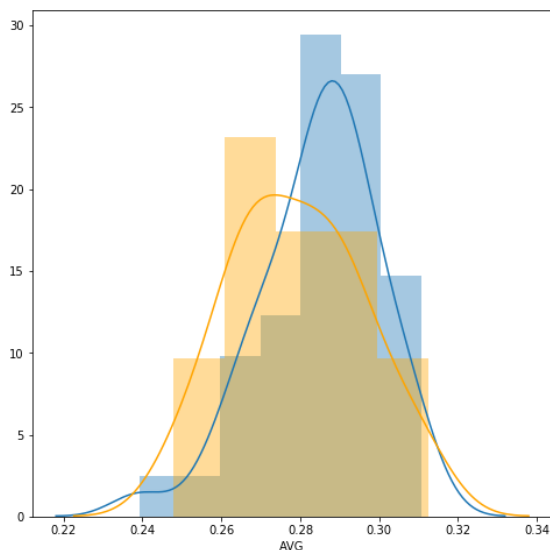
## 3.1.2. 데이터 정리

기록을 시즌별 팀 누적합으로 요약

```
home_train_x
```

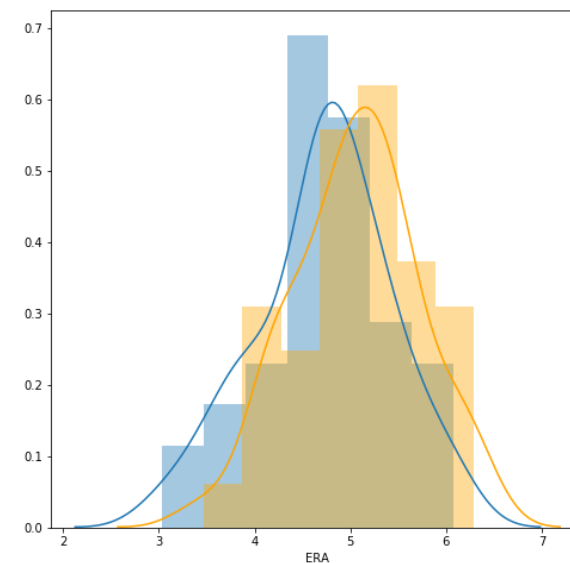
	PA	AB	RBI	RUN	HIT	H2	H3	HR	SB	CS	...	GD	H1	AVG	OPS	CQ	BABIP	BK
<b>16HH</b>	2223	1946	305	323	576	105	11	58	29	12	...	50	402	0.295992	0.822161	0.353043	0.340342	0.527273
<b>16HT</b>	2160	1889	295	316	541	105	11	60	39	24	...	36	365	0.286395	0.812171	0.347156	0.326544	0.514589
<b>16KT</b>	2246	1973	257	274	539	89	4	45	40	19	...	56	401	0.273188	0.736132	0.310477	0.311475	0.572222
<b>16LG</b>	2227	1931	296	319	563	87	19	35	57	27	...	29	422	0.291559	0.780115	0.337054	0.345324	0.543590
<b>16LT</b>	2166	1865	274	294	538	104	5	49	76	29	...	38	380	0.288472	0.803302	0.349382	0.343399	0.564792
...											...							...
<b>19NC</b>	2285	2014	288	308	576	102	11	61	34	18	...	50	402	0.285998	0.794945	0.334483	0.318688	0.531856
<b>19OB</b>	2241	1926	281	298	544	98	12	33	34	9	...	49	401	0.282451	0.761444	0.314815	0.319975	0.696697
<b>19SK</b>	2115	1865	263	272	496	82	6	56	43	17	...	31	352	0.265952	0.753643	0.320999	0.306407	0.505208
<b>19SS</b>	2220	1929	283	293	518	97	10	61	50	15	...	40	350	0.268533	0.771465	0.330971	0.306917	0.506173
<b>19WO</b>	2347	2067	316	335	590	114	25	42	46	9	...	46	409	0.285438	0.784043	0.332342	0.329723	0.555844

### 3.2.1. 홈-원정 간 편차 반영 변수 추가



홈  
원정

홈 대비  
원정 경기의  
분산은 큼



홈/원정에서의 실력 차이를 반영해야 하기 때문에 홈/원정간 편차를 반영해야 함

변수 추가

$\text{diff\_ER} = \text{HOME ER} - \text{AWAY ER}$

## 3.2.2. 변수별 T-test &amp; PCA

“변수의 유의성 판단하기 위하여 T-test 시행”

HOME

Variable	Coefficient
IB	1.381306
R	1.513901
LOB%	1.719611

⋮

AWAY

Variable	Coefficient
diff_ER	1.050310
IB	1.607985

⋮

Coefficient ≤ 2인  
변수 제거

“다중공선성 확인 후 PCA 시행”

HOME

VIF	features
978.99	SCORE%
902.16	K/BB
902.14	PA

⋮

AWAY

VIF	features
872.03	AVG
937.39	KK
535.82	BF

⋮

High VIF  
: 변수 간 다중 공선성“Feature Extraction:  
PCA를 통해  
다중 공선성 해결”



## 3.3.1. DATA SET 설정

✓ 2016~2019 DATA

Train set

Train set

X: 각 연도별 1~115 경기Y: 각 연도별 115~144 경기의 ER

팀별로 27~30의 잔여 경기  
: 마지막 30 잔여 경기를 예측하자!  
→ 115경기를 기준으로 나눔

✓ 2020 DATA

Test set

HOME

Test set

AWAY

X: 2020년 1~115경기

Y: 2020년 잔여 경기의 ER

: 예측해야 할 것

## 3.3.1. DATA SET 설정

 **STATIZ** 크롤링

홈/ 원정 팀별 투수 데이터  
Test set으로 활용  
(SF 변수 누락)



제공해주지 않은 2020 DATA는?

**KBO** 크롤링

홈/ 원정 나뉘어 있지 않아 각 경기 횟수의 비율로 SF 계산하여 추가



2020 DATA

Train set

Test set

HOME

Train set

Test set

AWAY

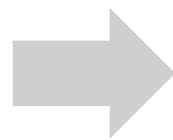
### 3.3.2. 홈/원정 데이터 분석 결과

#### 1) Sample weight 부여

$$\text{sample weight} = \frac{1}{\text{var}(\text{residual})^2}$$

#### 2) RMSE를 최소화하는 최적의 주성분 개수 (n\_component)

	n_component	RMSE
HOME	5	14.2756
AWAY	2	16.2811



HOME:0.886126

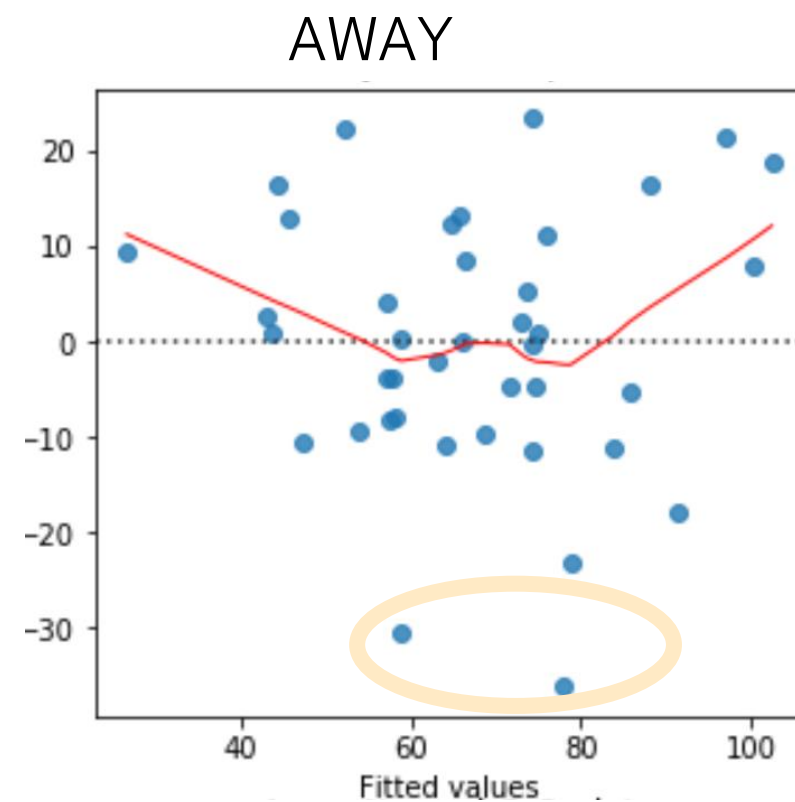
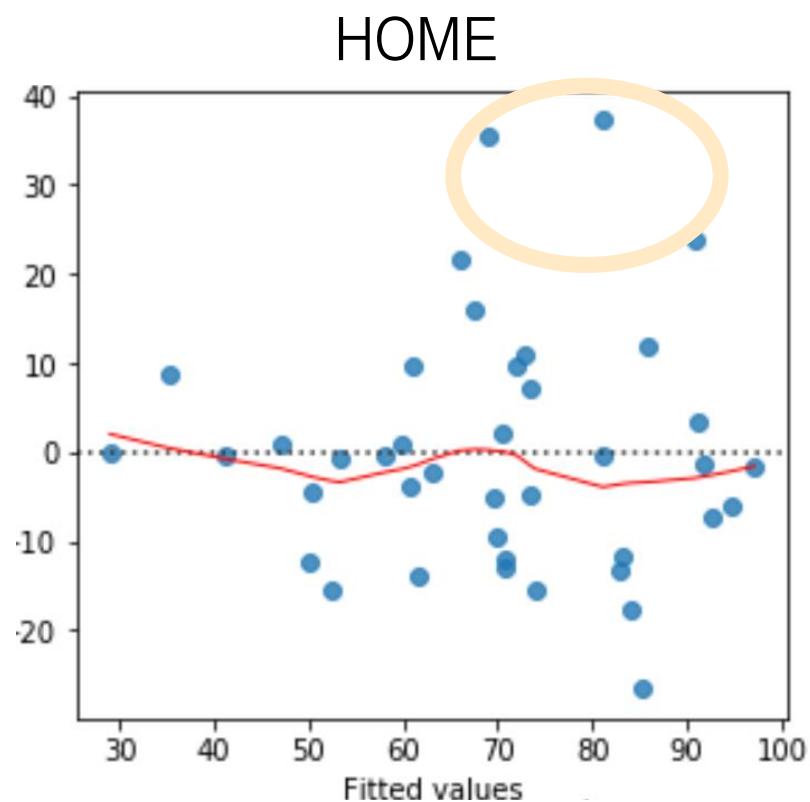
AWAY:0.767648

분산 설명력 확인

:홈/ 원정 모두 좋은 설명력 가짐

### 3.3.2. 홈/원정 데이터 분석 결과

#### 3) Residual analysis : Outlier 제거



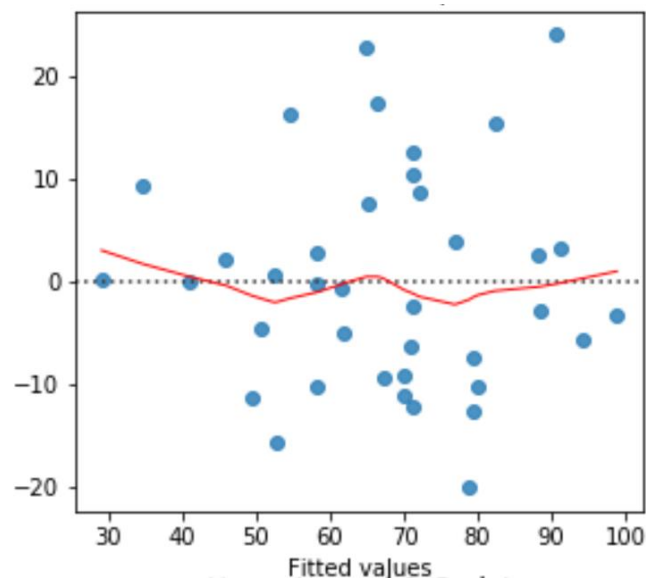
### 3.3.2. 홈/원정 데이터 분석 결과

#### 3) Residual analysis : Outlier 제거

Outlier

“그래프에서 눈에 띄는 Outlier 제거”

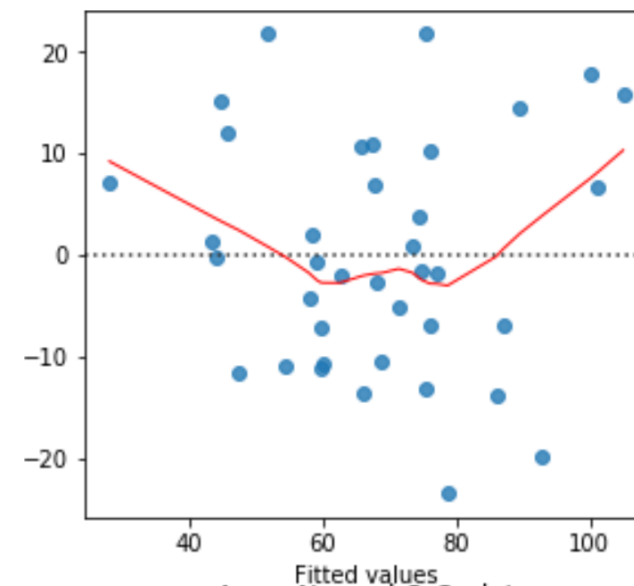
HOME



$R^2$   
0.9996  
→ 0.9935

RMSE  
13.516  
→ 10.578

AWAY



$R^2$   
0.9698  
→ 0.9758

RMSE  
13.512  
→ 11.396

### 3.3.2. 홈/원정 데이터 분석 결과

#### 3) Predict

잔여일 동안의 홈/ 원정 경기 수



홈/ 원정 예측값의 가중 평균

	T_ID	HOME_COUNT	AWAY_COUNT	HOME_PRED_ER	AWAY_PRED_ER	HOME_PRED_ERA	AWAY_PRED_ERA	ERA
0	HH	14	12	76.218974	53.465478	5.081265	4.158426	4.718169
1	HT	15	14	72.328573	58.960165	4.821905	4.211440	4.543094
2	KT	13	14	55.705758	52.293743	3.713717	3.237232	3.462594
3	LG	14	11	55.563351	48.480752	3.704223	4.113518	3.916043
4	LT	14	14	67.570185	54.540059	4.504679	3.636004	4.070341
5	NC	15	13	76.542658	39.140315	5.102844	3.010793	4.226949
6	OB	11	14	36.610094	62.527285	2.440673	3.275239	2.999258
7	SK	14	10	82.470412	46.143749	5.498027	4.306750	5.237112
8	SS	10	16	66.928132	68.865804	4.461875	2.869409	3.493361
9	WO	9	11	36.459026	48.645225	2.430602	2.653376	2.589686

### 3.3.2. 홈/원정 데이터 분석 결과

#### 3) Predict

최종 예측  
타율

	T_ID	ERA
9	WO	2.589686
6	OB	2.999258
2	KT	3.462594
8	SS	3.493361
3	LG	3.916043
4	LT	4.070341
5	NC	4.226949
1	HT	4.543094
0	HH	4.718169
7	SK	5.237112

## Chapter 4

# 승률

- 1) EDA & 데이터 전처리
- 2) Feature engineering
- 3) Data modeling



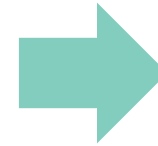
## DATA SET

팀	상대팀
HH	HT
	KT
	LG
	...
HT	HH
	KT
	LG
	LT
	...
...	

총 10팀  
상대팀으로 오는 팀 수=9팀  
총 90가지의 조합

X: 각 시즌 115경기 이후  
조합별 팀 AVG  
상대팀 AVG  
팀 ERA  
상대팀 ERA

Y: 각 시즌 115경기 이후  
조합별 팀 승률



## Predict

2020년도 115경기 이후  
조합별 승률

## 4.1. 분석 방향

## 과거 누적 데이터

G_ID	T_ID	VS_T_ID	TB_SC	WLS	team_gan_wp	team_wp	AVG	OBP	SLG	...	VS_ERA	VS_WHIP	VS_P_BABIP	VS_P_BK
20160507LTOB0	LT	OB	T	W	1.00	0.433333	0.281373	0.356153	0.417647	...	3.281374	1.331572	0.312757	2.336957
20160507SKSS0	SK	SS	T	W	0.00	0.600000	0.274354	0.344828	0.412525	...	5.216327	1.493878	0.323015	1.989474
20160508HHKT0	KT	HH	B	W	1.00	0.466667	0.277886	0.356342	0.400196	...	6.548638	1.828794	0.346253	1.310976
20160508HTWO0	WO	HT	B	W	0.75	0.551724	0.281190	0.356839	0.413628	...	4.280488	1.524390	0.320905	1.688073
20160508LTOB0	LT	OB	T	W	1.00	0.451613	0.280835	0.356784	0.417457	...	3.340561	1.339286	0.313830	2.319588
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
20200719LTSS0	LT	SS	T	W	0.25	0.475410	0.275152	0.344885	0.402433	...	4.460036	1.435169	0.292541	1.655462
20200719OBHT0	HT	OB	B	L	0.25	0.540984	0.276029	0.350258	0.422276	...	4.999407	1.520475	0.333333	2.019324
20200719OBHT0	OB	HT	T	W	0.75	0.587302	0.302491	0.369847	0.445285	...	4.292394	1.367207	0.318590	2.438503
20200719WOSK0	SK	WO	B	W	0.25	0.312500	0.243256	0.315240	0.360465	...	4.766551	1.402439	0.316826	2.371429
20200719WOSK0	WO	SK	T	L	0.75	0.584615	0.271788	0.357254	0.435310	...	4.874109	1.530285	0.314421	1.646840

일별 승/패 Binary Classification 시행

12	8
4	6

Confusion matrix

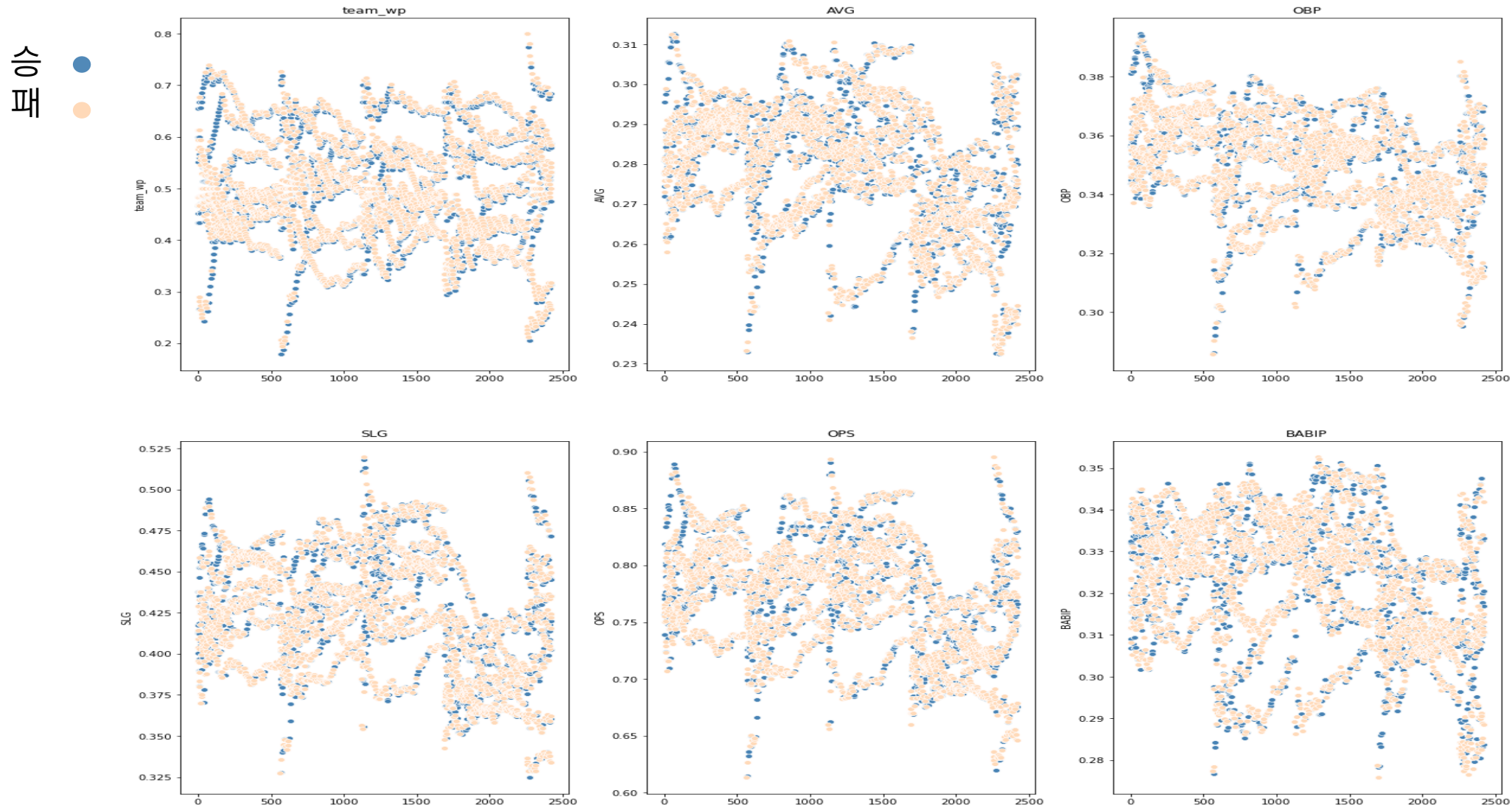
0.6

Accuracy score

임계값을 0.5(default)가 아닌  
다양한 값으로 조정해 보았음에도  
정확도가 좋지 않았음

## 4.1. 분석 방향

“과거 누적 데이터와 미래 승패는 관계가 없다”



## 4.1. 분석 방향

## 분석 방향 재설정

Before

After

데이터

과거 누적 데이터

예측 시점과 동일한 데이터

분석 방법

승/패 Binary Classification

승률 Regression

- 
- ① 오차를 감수하고, test set과 같은 시점의 데이터를 독립 변수로 활용하여 예측  
→ 앞 서 수행한 타율, 방어율 예측값
  - ① 승/패 여부: 상대팀에 크게 영향 받음 → 팀별 조합 생성하여 Feature engineering

## 4.2.1. Data set

## Train set

각 시즌 115경기 이후  
조합별 팀 AVG  
상대팀 AVG  
팀 ERA  
상대팀 ERA

## Test set

 **STATIZ** 크롤링 후

앞 서 홈/ 원정 데이터 Test set으로 활용하여  
타율/ 방어율 예측한 결과

## 4.2.2. 최적의 모델 및 Hyperparameter 설정 (Randomized search CV)

Randomized search cv 진행 결과 (n\_iter=20, k=4)

	Best Hyperparameter	Train MSE
Ridge regression	'alpha': 0.161	0.0769
Lasso regression	'alpha': 0.746	0.0831
Random forest	'n_estimators': 47, 'max_features': 'sqrt', 'max_depth': 5	0.0569
XGBoost	'booster'='gbtree', 'early_stopping_rounds'=20, 'n_estimators': 1, 'max_depth': 25, 'learning_rate': 0.329	0.0529

Train MSE가 비슷하게 낮은  
두 모델로 최종 예측

## 4.2.3. 분석 결과

	pred_wp_rf	pred_wp_wgb
T_ID		
HH	0.379882	0.495114
HT	0.422385	0.497180
KT	0.504173	0.518345
LG	0.446562	0.509468
LT	0.452328	0.509172
NC	0.579986	0.539819
OB	0.570334	0.474821
SK	0.401247	0.460703
SS	0.475578	0.500070
WO	0.549309	0.552092

올해 경기 양상을  
더 잘 반영하는  
Random forest로  
최종 예측



	T_ID	pred_wp_rf
0	HH	0.379882
1	HT	0.422385
2	KT	0.504173
3	LG	0.446562
4	LT	0.452328
5	NC	0.579986
6	OB	0.570334
7	SK	0.401247
8	SS	0.475578
9	WO	0.549309

Random forest & XGBoost 예측 결과

4.2.3. 분석 결과

최종 예측  
승률

	T_ID	pred_wp_rf
0	HH	0.379882
1	HT	0.422385
2	KT	0.504173
3	LG	0.446562
4	LT	0.452328
5	NC	0.579986
6	OB	0.570334
7	SK	0.401247
8	SS	0.475578
9	WO	0.549309



THANK YOU