

Class 2. SQLite3로 야구 스탯 직접 계산해보기

클래식 스탯과 세이버 스탯

스탯(stat): statistics 통계의 약자.

통계적인 수치로 나타낸 선수의 능력치.

스탯(stat) = 지표(indicator) = 지수(index)

1. 클래식 스탯 종류

타율, 홈런, 득점, 타점, 도루, 다승, 이닝, 평균자책점, 탈삼진, 볼넷, 출루율, 장타율 등

SABR(The Society for American Baseball Research)

- 1971년에 밥 데이비스 창시한 모임
- 야구를 통계학적/수학적으로 분석하는 방법론
- 빌 제임스: 세이버 메트릭스의 아버지

2. 세이버 스탯 종류

OPS(출루율+장타율), RC(출루율x장타율), BABIP(인플레이 타율), wOBA(가중 출루율), wRC+(조정 득점 생산력), FIP(수비 무관 평균자책점), DRS(디펜시스런세이브), WAR(승리 기여도), WHIP(이닝당 출루 허용), WPA(승리 확률 기여도) 등

참고1: 위키백과, 세이버메트릭스

참고2: <https://library.fangraphs.com/offense/offensive-statistics-list/>

KBO 홈페이지 - 기록실

기록실

선수기록 팀기록 기록용어 ?

타자 투수 수비 주루

2021 KBO 정규시즌 팀 선택 포지션 선택 경기상황별1 경기상황별2

타자기록

기본기록 세부기록 기록보기

순위	선수명	팀명	AVG	G	PA	AB	R	H	2B	3B	HR	TB	RBI	SAC	SF
1	이정후	키움	0.358	121	535	455	76	163	42	6	6	235	82	0	9
2	전준우	롯데	0.348	141	605	538	88	187	43	0	7	251	91	1	8
3	강백호	KT	0.347	138	608	502	73	174	38	1	16	262	100	0	8
4	홍창기	LG	0.327	141	637	513	101	168	26	2	4	210	51	0	2
5	박건우	두산	0.327	123	512	447	78	146	31	2	5	196	60	0	6
6	손아섭	롯데	0.323	136	598	533	88	172	29	2	3	214	57	3	0
7	양의지	NC	0.323	137	553	468	80	151	28	2	29	270	109	0	10
8	페르난데스	두산	0.319	138	603	527	72	168	24	0	15	237	80	0	6
9	김선빈	KIA	0.307	130	564	501	55	154	32	0	5	201	67	1	5
10	안치훈	롯데	0.307	116	484	417	58	128	30	2	10	192	82	5	9

스탯티즈 - 기록실

시즌기록실 시즌기록실 통산기록실 팀기록실 특별기록실 연도별 상수 WAR Special

시즌기록실

2021 연도 시작 끝 팀:전체 포지션 정규 규정 상황 옵션

[자동 : 규정 100% 이상]

공지 투수 WAR 계산 중 상대팀 공격력을 반영하는 과정에서 SSG 팀 관련 오류가 있음을 확인하고 현재 수정 완료 했습니다. 혼란을 드린점 진심으로 사과 드립니다.

기본 확장 가치 클러치 타석 타구1 타구2 파워 팀배팅1 팀배팅2 도루 주루 구종가치 구종구사

순	이름	팀	생산력+	타석	HR%	BB%	K%	BB/K	IsoP	IsoD	BABIP	Spd	PSN	타격 생산력				타격 생산력 (파크팩터 조정)				
			wRC+											wOBA	wRC	wRC ₂₇	wRAA	wOBA	wRC	wRC ₂₇	wRAA	wRC+
1	강백호	21 K 1B	164.3	608	2.63	16.1	13.8	1.17	.175	.101	.385	3.7	12.31	.442	125.1	9.71	50.1	.438	123.3	9.57	48.3	164.3
2	이정후	21 K CF	164.0	534	1.12	11.6	6.9	1.68	.159	.079	.374	5.8	7.50	.440	108.9	9.46	43.0	.438	108.1	9.39	42.2	164.0
3	양의지	21 N DH	162.7	553	5.24	11.6	10.5	1.10	.254	.086	.312	3.1	3.74	.439	112.5	8.93	44.2	.436	111.1	8.82	42.8	162.7
4	최정	21 S 3B	153.8	547	6.40	15.0	18.3	0.82	.286	.132	.270	3.6	13.02	.429	106.8	8.48	39.2	.422	103.9	8.25	36.3	153.8
5	홍창기	21 L CF	152.4	637	0.63	16.8	14.6	1.15	.082	.128	.392	5.3	6.82	.417	118.1	8.83	39.5	.420	119.9	8.96	41.2	152.4
6	김재환	21 두 LF	144.3	553	4.52	14.3	22.6	0.63	.220	.109	.314	3.5	3.70	.395	92.7	7.05	24.4	.408	98.5	7.49	30.2	144.3
7	전준우	21 롯데 LF	139.7	605	1.16	8.8	11.4	0.77	.119	.058	.383	3.6	6.46	.406	106.8	7.73	32.1	.401	104.3	7.55	29.7	139.7
8	페르난데스	21 두 DH	139.7	603	2.49	10.6	5.6	1.88	.131	.076	.316	1.3	0.00	.393	100.1	6.89	25.6	.401	104.0	7.16	29.5	139.7
9	박건우	21 두 RF	137.7	512	0.98	9.4	13.5	0.70	.112	.074	.372	4.9	7.06	.393	85.0	7.13	21.8	.398	87.0	7.30	23.8	137.7
10	주신수	21 S DH	136.5	571	3.68	18.0	21.0	0.86	.184	.147	.311	4.6	22.83	.400	97.9	7.47	27.4	.396	96.2	7.34	25.7	136.5

KBO <https://www.koreabaseball.com/>

스탯티즈 <http://www.statiz.co.kr/main.php>

야구 스탯 sqlite3로 계산해보기

1. 타율

안타/타수 = 안타/(타석수 - 볼넷 - 몸에맞는공 - 희생번트 - 희생플라이 - 타격방해 - 주루방해)

예제1. 추신수 선수 2019년 타율을 구해보자

Step1. 우선 추신수 선수의 playerId 확인하기

- 선수들의 playerId는 people 테이블에서 확인 가능
- 성이 Choo인 선수를 찾아보자

```
SELECT * FROM people WHERE nameLast = 'Choo';
```

```
choosh01|1982|7|13|South Korea|Busan|Busan|||||
|Shin-Soo|Choo|Shin-Soo|210|71|L|L|2005-04-21|20
19-09-29|choos001|choosh01|1982-07-13|2005-04-21
|2019-09-29|
```

.schema people

```
sqlite> .schema people
CREATE TABLE IF NOT EXISTS "people" (
  "playerID" VARCHAR(9) NOT NULL,
  "birthYear" INTEGER NULL,
  "birthMonth" INTEGER NULL,
  "birthDay" INTEGER NULL,
  "birthCountry" VARCHAR(255) NULL,
  "birthState" VARCHAR(255) NULL,
  "birthCity" VARCHAR(255) NULL,
  "deathYear" INTEGER NULL,
  "deathMonth" INTEGER NULL,
  "deathDay" INTEGER NULL,
  "deathCountry" VARCHAR(255) NULL,
  "deathState" VARCHAR(255) NULL,
  "deathCity" VARCHAR(255) NULL,
  "nameFirst" VARCHAR(255) NULL,
  "nameLast" VARCHAR(255) NULL,
  "nameGiven" VARCHAR(255) NULL,
  "weight" INTEGER NULL,
  "height" INTEGER NULL,
  "bats" VARCHAR(255) NULL,
  "throws" VARCHAR(255) NULL,
  "debut" VARCHAR(255) NULL,
  "finalGame" VARCHAR(255) NULL,
  "retroID" VARCHAR(255) NULL,
  "bbrefID" VARCHAR(255) NULL,
  "birth_date" DATE NULL,
  "debut_date" DATE NULL,
  "finalgame_date" DATE NULL,
  "death_date" DATE NULL,
  PRIMARY KEY ("playerID")
);
```

야구 스탯 sqlite3로 계산해보기

1. 타율

안타/타수 = 안타/(타석수 - 볼넷 - 몸에맞는공 - 희생번트 - 희생플라이 - 타격방해 - 주루방해)

Step2. 추신수 선수의 2019년 타율 계산해서 조회

```
SELECT playerID, CAST(H AS REAL)/AB AS AVG FROM batting WHERE yearID = 2019 and playerID = 'choosh01';
```

playerID	AVG
choosh01	0.264653641207815

2014	31	TEX	AL	123	529	455	58	110	19	1	13	40	3	4	58	131	.242	.340	.374	.714
2015	32	TEX	AL	149	653	555	94	153	32	3	22	82	4	2	76	147	.276	.375	.463	.838
2016	33	TEX	AL	48	210	178	27	43	7	0	7	17	6	3	25	46	.242	.357	.399	.756
2017	34	TEX	AL	149	636	544	96	142	20	1	22	78	12	3	77	134	.261	.357	.423	.780
2018	35	TEX	AL	146	665	560	83	148	30	1	21	62	6	1	92	156	.261	.377	.434	.810
2019	36	TEX	AL	151	660	563	93	149	31	2	24	61	15	1	78	165	.265	.371	.455	.826
2020	37	TEX	AL	33	127	110	13	26	3	0	5	15	6	2	13	33	.236	.323	.400	.723

cast() 함수

형변환(type casting)에 사용되는 함수. 형변환이란 자료형을 다른 형태로 변경하는 것을 의미.

ex) SELECT CAST(1.25 AS INT);

```
sqlite> SELECT CAST(1.25 AS INT);
1
sqlite> SELECT CAST(1 AS REAL);
1.0
```

AS 별칭

Alias의 약어. 별명이라는 뜻.
컬럼의 이름을 별도의 별칭으로 지정하고 싶을 때 사용.

ex) SELECT CAST(H AS REAL)/AB AS AVG;

야구 스탯 sqlite3로 계산해보기

2. 장타율 (Slugging Percentage, SLG)

(1루타 + 2*2루타 + 3*3루타 + 4*홈런) / 타수

예제2. 추신수 선수 2019년 장타율을 구해보자

```
SELECT playerID, CAST((((H - "2B" - "3B" - HR) + 2*"2B" + 3*"3B" + 4*HR) AS REAL)/AB AS SLG FROM batting WHERE yearID = 2019 and playerID = 'choosh01';
```

playerID	SLG
choosh01	0.454706927175844

2014	31	TEX	AL	123	529	455	58	110	19	1	13	40	3	4	58	131	.242	.340	.374	.714
2015	32	TEX	AL	149	653	555	94	153	32	3	22	82	4	2	76	147	.276	.375	.463	.838
2016	33	TEX	AL	48	210	178	27	43	7	0	7	17	6	3	25	46	.242	.357	.399	.756
2017	34	TEX	AL	149	636	544	96	142	20	1	22	78	12	3	77	134	.261	.357	.423	.780
2018 *	35	TEX	AL	146	665	560	83	148	30	1	21	62	6	1	92	156	.264	.377	.434	.810
2019	36	TEX	AL	151	660	563	93	149	31	2	24	61	15	1	78	165	.265	.371	.455	.826
2020	37	TEX	AL	33	127	110	13	26	3	0	5	15	6	2	13	33	.236	.323	.400	.723

3. 출루율 (On Base Percentage, OBP)

(안타 + 볼넷 + 몸에맞은공) / (타수 + 볼넷 + 몸에맞은공 + 희생플라이)

예제3. 추신수 선수 2019년 출루율을 구해보자

```
SELECT playerID, CAST((H + BB + HBP) AS REAL)/(AB + BB + HBP + SF) AS OBP FROM batting WHERE yearID = 2019 and playerID = 'choosh01';
```

playerID	OBP
choosh01	0.371212121212121

2014	31	TEX	AL	123	529	455	58	110	19	1	13	40	3	4	58	131	.242	.340	.374	.714
2015	32	TEX	AL	149	653	555	94	153	32	3	22	82	4	2	76	147	.276	.375	.463	.838
2016	33	TEX	AL	48	210	178	27	43	7	0	7	17	6	3	25	46	.242	.357	.399	.756
2017	34	TEX	AL	149	636	544	96	142	20	1	22	78	12	3	77	134	.261	.357	.423	.780
2018 *	35	TEX	AL	146	665	560	83	148	30	1	21	62	6	1	92	156	.264	.377	.434	.810
2019	36	TEX	AL	151	660	563	93	149	31	2	24	61	15	1	78	165	.265	.371	.455	.826
2020	37	TEX	AL	33	127	110	13	26	3	0	5	15	6	2	13	33	.236	.323	.400	.723

야구 스탯 sqlite3로 계산해보기

4. OPS (On-base Plus Slugging)

장타율 + 출루율

예제4-1. 추신수의 2019년 OPS를 구해보자

```
SELECT playerID, CAST((((H - "2B" - "3B" - HR) + 2*"2B" + 3*"3B" + 4*HR) AS REAL)/AB + CAST((H + BB + HBP) AS REAL)/(AB + BB + HBP + SF) AS OPS FROM batting WHERE yearID = 2019 and playerID = 'choosh01';
```

playerID	OPS
choosh01	0.825919048387965

2014	31	TEX	AL	123	529	455	58	110	19	1	13	40	3	4	58	131	.242	.340	.374	.714
2015	32	TEX	AL	149	653	555	94	153	32	3	22	82	4	2	76	147	.276	.375	.463	.838
2016	33	TEX	AL	48	210	178	27	43	7	0	7	17	6	3	25	46	.242	.357	.399	.756
2017	34	TEX	AL	149	636	544	96	142	20	1	22	78	12	3	77	134	.261	.357	.423	.780
2018 ★	35	TEX	AL	146	665	560	83	148	30	1	21	62	6	1	92	156	.264	.377	.434	.810
2019	36	TEX	AL	151	660	563	93	149	31	2	24	61	15	1	78	165	.265	.371	.455	.826
2020	37	TEX	AL	33	127	110	13	26	3	0	5	15	6	2	13	33	.236	.323	.400	.723

잘 구했지만 OPS를 구하는 식이 너무 길어서 복잡함.
장타율과 출루율을 각각 조회한 것을 더하는 방법은 없을까?

야구 스탯 sqlite3로 계산해보기

4. OPS (On-base Plus Slugging)

장타율 + 출루율

예제4-2. 추신수의 2019년 OPS를 구해보자

서브쿼리(subquery) 활용하기

```
SELECT playerID, CAST((((H - "2B" - "3B" - HR) + 2*"2B" + 3*"3B" + 4*HR) AS REAL)/AB AS SLG, CAST((H + BB + HBP) AS REAL)/(AB + BB + HBP + SF) AS OBP FROM batting WHERE yearID = 2019 and playerID = 'choosh01';
```

playerID	SLG	OBP
choosh01	0.454706927175844	0.371212121212121

서브쿼리

메인쿼리

```
SELECT playerID, (SLG + OBP) AS OPS FROM (위에서 조회한 결과 테이블);
```

```
SELECT playerID, (SLG + OBP) AS OPS FROM (SELECT playerID, CAST((((H - "2B" - "3B" - HR) + 2*"2B" + 3*"3B" + 4*HR) AS REAL)/AB AS SLG, CAST((H + BB + HBP) AS REAL)/(AB + BB + HBP + SF) AS OBP FROM batting WHERE yearID = 2019 and playerID = 'choosh01');
```

playerID	OPS
choosh01	0.825919048387965

2014	31	TEX	AL	123	529	455	58	110	19	1	13	40	3	4	58	131	.242	.340	.374	.714
2015	32	TEX	AL	149	653	555	94	153	32	3	22	82	4	2	76	147	.276	.375	.463	.838
2016	33	TEX	AL	48	210	178	27	43	7	0	7	17	6	3	25	46	.242	.357	.399	.756
2017	34	TEX	AL	149	636	544	96	142	20	1	22	78	12	3	77	134	.261	.357	.423	.780
2018	35	TEX	AL	146	665	560	83	148	30	1	21	62	6	1	92	156	.264	.377	.434	.810
2019	36	TEX	AL	151	660	563	93	149	31	2	24	61	15	1	78	165	.265	.371	.453	.826
2020	37	TEX	AL	33	127	110	13	26	3	0	5	15	6	2	13	33	.236	.323	.400	.723

야구 스탯 sqlite3로 계산해보기

5. BABIP(batting average on balls in play): 인플레이 타구의 안타 비율
(안타 - 홈런)/(타수 - 삼진 - 홈런 + 희생플라이)

예제5-1. 추신수의 2019년 BABIP을 구해보자

```
SELECT playerID, CAST((H - HR) AS REAL)/(AB - SO - HR + SF) AS BABIP FROM batting
WHERE yearID = 2019 and playerID = 'choosh01';
```

playerID	BABIP
choosh01	0.3333333333333333

2017	34	TEX	AL	.349	103	.305	.162	3.5%	21.1%	12.1%
2018	35	TEX	AL	.368	121	.330	.170	3.2%	23.5%	13.8%
2019	36	TEX	AL	.371	116	.333	.190	3.6%	25.0%	11.8%
2020	37	TEX	AL	.321	97	.284	.164	3.9%	26.0%	10.2%



브로스 맥브라켄
"인플레이 타구가 안타가 될지 범타가 될지는 투수가 통제할 수 없다"

참고자료1: <https://namu.wiki/w/BABIP>

야구 스탯 sqlite3로 계산해보기

5. BABIP(batting average on balls in play): 인플레이 타구의 안타 비율

(안타 - 홈런)/(타수 - 삼진 - 홈런 + 희생플라이)

* 타수 = 타석수 - 볼넷 - 몸에맞은공 - 희생번트 - 희생플라이 - 타격방해 - 주루방해

예제5-2. 류현진의 2019년 BABIP을 구해보자

```
SELECT playerID, CAST((H - HR) AS REAL)/((BFP - BB - HBP - SH - SF) - SO - HR + SF)
AS BABIP FROM pitching WHERE yearID = 2019 and playerID = 'ryuhy01';
```

playerID	BABIP
ryuhy01	0.282051282051282

					Batting Against				
Year	Age	Tm	Lg	IP	BA	OBP	SLG	OPS	BAbip
2013	26	LAD	NL	192.0	.252	.299	.361	.660	.300
2014	27	LAD	NL	152.0	.257	.294	.364	.658	.323
2016	29	LAD	NL	4.2	.364	.417	.727	1.144	.412
2017	30	LAD	NL	126.2	.263	.330	.462	.792	.303
2018	31	LAD	NL	82.1	.221	.260	.362	.622	.282
2019	32	LAD	NL	182.2	.234	.263	.359	.622	.282
2020	33	TOR	AL	67.0	.234	.285	.352	.636	.303
2021	34	TOR	AL	169.0	.258	.299	.435	.733	.296

야구 스탯 sqlite3로 계산해보기

6. WHIP(이닝당 안타 및 볼넷 허용률) (피안타 + 볼넷)/이닝

예제6. 클레이튼 커쇼의 2019년 WHIP를 구해보자

Step1. 우선 커쇼의 playerId 확인하기

```
SELECT * FROM people WHERE nameLast = 'Kershaw';
```

playerID	birthYear	birthMonth	birthDay	birthCountry	birthState	birthCity	deathYear	deathMonth	deathDay	deathCountry	deathState	deathCity	nameFirst	nameLast	nameGiven	weight	height	bats	throws	debut	finalGame
retroID	bbrefID	birth_date	debut_date	finalgame_date	death_date																
kershcl01		1988	3	19	USA	TX	Dallas						Clayton	Kershaw	Clayton Edward	226	76	L	L	2008-05-25	2019-09-29
kersc001	kershcl01	1988-03-19	2008-05-25	2019-09-29																	

Step2. 커쇼의 WHIP 조회하기

```
SELECT playerId, CAST((H + BB) AS REAL)/(CAST(IPouts AS REAL)/3) AS WHIP FROM pitching WHERE yearID = 2019 and playerId = 'kershcl01';
```

playerID	WHIP
kershcl01	1.04299065420561

2013	★	25	LAD	NL	16	9	.640	1.83	33	33	0	3	2	0	236.0	164	55	48	11	52	2	232	3	2	12	908	194	2.39	0.915	6.3	0.4	2.0
2014	★	26	LAD	NL	21	3	.875	1.77	27	27	0	6	2	0	198.1	139	42	39	9	31	0	239	2	2	7	749	197	1.81	0.857	6.3	0.4	1.4
2015	★	27	LAD	NL	16	7	.696	2.13	33	33	0	4	3	0	232.2	163	62	55	15	42	1	301	5	3	9	890	173	1.99	0.881	6.3	0.6	1.6
2016	★	28	LAD	NL	12	4	.750	1.69	21	21	0	3	3	0	149.0	97	31	28	8	11	1	172	2	3	5	544	237	1.80	0.725	5.9	0.5	0.7
2017	★	29	LAD	NL	18	4	.818	2.31	27	27	0	1	0	0	175.0	136	49	45	23	30	0	202	0	2	4	679	179	3.07	0.949	7.0	1.2	1.5
2018		30	LAD	NL	9	5	.643	2.73	26	26	0	0	0	0	161.1	139	55	49	17	29	0	155	2	0	10	650	142	3.18	1.041	7.8	0.9	1.6
2019	★	31	LAD	NL	16	5	.762	3.03	29	28	0	0	0	0	178.1	145	63	60	28	41	0	189	2	1	7	706	137	3.86	1.043	.3	1.4	2.1
2020		32	LAD	NL	6	2	.750	2.16	10	10	0	0	0	0	58.1	41	18	14	8	8	0	62	1	0	0	221	202	3.31	0.840	6.3	1.2	1.2
2021		33	LAD	NL	10	8	.556	3.55	22	22	0	0	0	0	121.2	103	51	48	15	21	0	144	3	1	7	488	115	3.00	1.019	7.6	1.1	1.6

과제#2

관심 있는 선수들의 스탯 5개 계산해보기(타율, OPS, BABIP, WHIP, FIP, wOBA 등)

강의자료에서 공식을 소개하지 않은 스탯의 경우 공식을 직접 찾아보고 쿼리로 구해볼 것!

스탯 계산 공식과 쿼리 및 조회 결과 캡처 화면을 word로 정리해서 kyohoonsim@gmail.com 으로 보내주세요~!

문서 제목 양식:

KUSF데이터분석_과제2_이름.docx

ex) KUSF데이터분석_과제2_심교훈.docx