14U Player Values

June 8, 2023

1 Zachary Inn

[1]:
```
import warnings
warnings.filterwarnings('ignore')

import pandas as pd
import numpy as np
from plotnine import *

from sklearn.linear_model import LinearRegression # Linear Regression Model
from sklearn.preprocessing import StandardScaler #Z-score variables
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score␣
 ↪#model evaluation
from sklearn.model_selection import train_test_split # simple TT split cv
from heapq import nsmallest
```

2 Import Dataset

[2]:
```
data = pd.read_csv("https://raw.githubusercontent.com/ZacharyInn/
 ↪WreckingYardData/main/CSVFiles/
 ↪Wrecking%20Crew%202027%2014U%20Fall%202022%20Stats.csv")

data
```

[2]:

	Number	Last	First	GP	PA	AB	AVG	OBP	OPS	SLG	\
0	00	Cha	Bladen	43	116	78	0.410	0.578	1.065	0.487	
1	2	Mertz	Will	41	108	74	0.216	0.454	0.751	0.297	
2	3	Inn	Nathan	39	102	76	0.263	0.431	0.773	0.342	
3	7	Hong	Jonas	19	49	41	0.293	0.408	0.823	0.415	
4	8	Fang	Aidan	35	82	48	0.250	0.561	0.832	0.271	
5	11	Bryson	Tyler	40	99	77	0.260	0.414	0.726	0.312	
6	20	Blankenship	Sean	37	83	71	0.296	0.398	0.792	0.394	
7	27	Lie	Warren	36	84	69	0.261	0.393	0.668	0.275	
8	33	Robinson	Kolby	32	61	51	0.176	0.311	0.488	0.176	
9	34	Nussbaum	Peter	35	89	71	0.380	0.483	1.145	0.662	
10	Totals	NaN	NaN	43	1111	837	0.277	0.445	0.803	0.357	

	…	DP	TP	INN	PB	SB.2	SB-ATT	CS.2	CS%	PIK.2	CI.1
0	…	1	1	0.0	0	0	0-0	0	0.00	0	0
1	…	5	0	0.0	0	0	0-0	0	0.00	0	0
2	…	2	1	0.0	0	0	0-0	0	0.00	0	0
3	…	0	0	0.0	0	0	0-0	0	0.00	0	0
4	…	3	0	109.2	3	84	84-96	12	12.50	2	0
5	…	2	0	0.0	0	0	0-0	0	0.00	0	0
6	…	4	0	0.0	0	0	0-0	0	0.00	0	0
7	…	1	0	0.0	0	0	0-0	0	0.00	0	0
8	…	0	0	0.0	0	0	0-0	0	0.00	0	0
9	…	5	1	3.0	2	2	2-3	1	33.33	0	0
10	…	23	1	212.1	14	125	125-146	21	14.38	2	0

[11 rows x 157 columns]

3 Introduction

One of the most common uses of advanced statistics in baseball to find a player's true value to a team that takes into account all aspects of their game. An example of this would be defense first players who are usually undervalued because they don't hit as well. This player may be playing defense so well that it makes up for their lack of production at the plate since they are saving runs from being scored against them. In this situation, taking this defensively minded player out for a worse defender but a better bat might be the wrong option as their bat may not make up for their glove.

In this report, I wanted to accomplish two things: look at statistics to see if I can extract something out of every player's value and create a batting lineup that is based on a sabermetric take of the game. I will be looking at a variety of stats on GameChanger from the Fall of 2022 in order to find hidden gems within these players. Also I will create a batting lineup because I find it interesting to structure lineups and want to see how it differs from this team's typical lineup.

The players in this data set all play for the 14U age division and I only included those who had at least one plate appearance per game (43 games). I also manually calculated a couple of other useful stats that GameChanger doesn't have below.

4 Calculating wOBA (Weighted On-Base Average)

[3]:
```
data["wOBA"] = ((data["BB"].astype(float) * .69) + (data["HBP"].astype(float) *␣
 ↪.72) + data["1B"].astype(float) * .89 + data["2B"].astype(float) * 1.27 +␣
 ↪data["3B"].astype(float) * 1.62 + data["HR"].astype(float) * 2.1) /␣
 ↪(data["AB"].astype(float) + data["BB"].astype(float) + data["SF"].
 ↪astype(float) + data["HBP"].astype(float))
data
```

[3]:

	Number	Last	First	GP	PA	AB	AVG	OBP	OPS	SLG	\
0	00	Cha	Bladen	43	116	78	0.410	0.578	1.065	0.487	
1	2	Mertz	Will	41	108	74	0.216	0.454	0.751	0.297	
2	3	Inn	Nathan	39	102	76	0.263	0.431	0.773	0.342	
3	7	Hong	Jonas	19	49	41	0.293	0.408	0.823	0.415	
4	8	Fang	Aidan	35	82	48	0.250	0.561	0.832	0.271	
5	11	Bryson	Tyler	40	99	77	0.260	0.414	0.726	0.312	
6	20	Blankenship	Sean	37	83	71	0.296	0.398	0.792	0.394	
7	27	Lie	Warren	36	84	69	0.261	0.393	0.668	0.275	
8	33	Robinson	Kolby	32	61	51	0.176	0.311	0.488	0.176	
9	34	Nussbaum	Peter	35	89	71	0.380	0.483	1.145	0.662	
10	Totals	NaN	NaN	43	1111	837	0.277	0.445	0.803	0.357	

	…	TP	INN	PB	SB.2	SB-ATT	CS.2	CS%	PIK.2	CI.1	wOBA
0	…	1	0.0	0	0	0-0	0	0.00	0	0	0.474655
1	…	0	0.0	0	0	0-0	0	0.00	0	0	0.366389
2	…	1	0.0	0	0	0-0	0	0.00	0	0	0.361863
3	…	0	0.0	0	0	0-0	0	0.00	0	0	0.370000
4	…	0	109.2	3	84	84-96	12	12.50	2	0	0.424268
5	…	0	0.0	0	0	0-0	0	0.00	0	0	0.343030
6	…	0	0.0	0	0	0-0	0	0.00	0	0	0.358434
7	…	0	0.0	0	0	0-0	0	0.00	0	0	0.319167
8	…	0	0.0	0	0	0-0	0	0.00	0	0	0.244426
9	…	1	3.0	2	2	2-3	1	33.33	0	0	0.480337
10	…	1	212.1	14	125	125-146	21	14.38	2	0	0.373670

[11 rows x 158 columns]

According to MLB.com, "wOBA is a version of on-base percentage that accounts for how a player reached base – instead of simply considering whether a player reached base." In short, reaching base via a double is more valuable than reaching base with a walk since a double usually can score runs more often than a walk.

5 Calculating Runs Created

[4]:
```
data["RC"] = data["TB"].astype(float) * (data["BB"].astype(float) + data["H"].
↪astype(float)) / (data["BB"].astype(float) + data["AB"].astype(float))
data
```

[4]:

	Number	Last	First	GP	PA	AB	AVG	OBP	OPS	SLG	\
0	00	Cha	Bladen	43	116	78	0.410	0.578	1.065	0.487	
1	2	Mertz	Will	41	108	74	0.216	0.454	0.751	0.297	
2	3	Inn	Nathan	39	102	76	0.263	0.431	0.773	0.342	
3	7	Hong	Jonas	19	49	41	0.293	0.408	0.823	0.415	
4	8	Fang	Aidan	35	82	48	0.250	0.561	0.832	0.271	
5	11	Bryson	Tyler	40	99	77	0.260	0.414	0.726	0.312	

```
6       20  Blankenship    Sean  37    83   71  0.296  0.398  0.792  0.394
7       27          Lie  Warren  36    84   69  0.261  0.393  0.668  0.275
8       33     Robinson   Kolby  32    61   51  0.176  0.311  0.488  0.176
9       34     Nussbaum   Peter  35    89   71  0.380  0.483  1.145  0.662
10  Totals          NaN     NaN  43  1111  837  0.277  0.445  0.803  0.357

    …    INN  PB  SB.2   SB-ATT  CS.2    CS%  PIK.2  CI.1      wOBA  \
0   …    0.0   0     0      0-0     0   0.00      0     0  0.474655
1   …    0.0   0     0      0-0     0   0.00      0     0  0.366389
2   …    0.0   0     0      0-0     0   0.00      0     0  0.361863
3   …    0.0   0     0      0-0     0   0.00      0     0  0.370000
4   …  109.2   3    84    84-96    12  12.50      2     0  0.424268
5   …    0.0   0     0      0-0     0   0.00      0     0  0.343030
6   …    0.0   0     0      0-0     0   0.00      0     0  0.358434
7   …    0.0   0     0      0-0     0   0.00      0     0  0.319167
8   …    0.0   0     0      0-0     0   0.00      0     0  0.244426
9   …    3.0   2     2      2-3     1  33.33      0     0  0.480337
10  …  212.1  14   125  125-146    21  14.38      2     0  0.373670

            RC
0    21.663551
1     9.240000
2    10.000000
3     6.729167
4     6.589041
5     9.130435
6    10.278481
7     7.182927
8     2.803279
9    22.953488
10  125.728927

[11 rows x 159 columns]
```

According to MLB.com, "Runs Created estimates a player's offensive contribution in terms of total runs." It combines a player's ability to get on base and hit for extra bases and divides it by all of their opportunities.

6 Calculating ISO (Isolated Power)

[5]:
```
data["ISO"] = data["SLG"].astype(float) - data["AVG"].astype(float)
data
```

[5]:
```
    Number         Last   First  GP    PA   AB    AVG    OBP    OPS    SLG  \
0       00          Cha  Bladen  43   116   78  0.410  0.578  1.065  0.487
1        2       Mertz    Will  41   108   74  0.216  0.454  0.751  0.297
```

2	3	Inn	Nathan	39	102	76	0.263	0.431	0.773	0.342
3	7	Hong	Jonas	19	49	41	0.293	0.408	0.823	0.415
4	8	Fang	Aidan	35	82	48	0.250	0.561	0.832	0.271
5	11	Bryson	Tyler	40	99	77	0.260	0.414	0.726	0.312
6	20	Blankenship	Sean	37	83	71	0.296	0.398	0.792	0.394
7	27	Lie	Warren	36	84	69	0.261	0.393	0.668	0.275
8	33	Robinson	Kolby	32	61	51	0.176	0.311	0.488	0.176
9	34	Nussbaum	Peter	35	89	71	0.380	0.483	1.145	0.662
10	Totals	NaN	NaN	43	1111	837	0.277	0.445	0.803	0.357

	…	PB	SB.2	SB-ATT	CS.2	CS%	PIK.2	CI.1	wOBA	RC	\
0	…	0	0	0-0	0	0.00	0	0	0.474655	21.663551	
1	…	0	0	0-0	0	0.00	0	0	0.366389	9.240000	
2	…	0	0	0-0	0	0.00	0	0	0.361863	10.000000	
3	…	0	0	0-0	0	0.00	0	0	0.370000	6.729167	
4	…	3	84	84-96	12	12.50	2	0	0.424268	6.589041	
5	…	0	0	0-0	0	0.00	0	0	0.343030	9.130435	
6	…	0	0	0-0	0	0.00	0	0	0.358434	10.278481	
7	…	0	0	0-0	0	0.00	0	0	0.319167	7.182927	
8	…	0	0	0-0	0	0.00	0	0	0.244426	2.803279	
9	…	2	2	2-3	1	33.33	0	0	0.480337	22.953488	
10	…	14	125	125-146	21	14.38	2	0	0.373670	125.728927	

	ISO
0	0.077
1	0.081
2	0.079
3	0.122
4	0.021
5	0.052
6	0.098
7	0.014
8	0.000
9	0.282
10	0.080

[11 rows x 160 columns]

According to MLB.com, "ISO measures the raw power of a hitter by taking only extra-base hits – and the type of extra-base hit – into account."

7 Interesting stats from each player, hitting wise

Cha - Cha is your dream lead-off hitter, not only does he have the highest OBP on the team (.578), but he also slugs (.487), has a 61.2% quality AB rate, walks twice as much as he strikes out, and has the highest amount of total bases (38).

Mertz - In a classic lineup, Mertz would probably be seen as a great leadoff hitter. He leads the

team with 26 stolen bases, leads the team with runs scored with 40, leads the team with 35 plate appearances seeing 6 or more pitches, and has the second most walks with 26. If Mertz excels at getting on base and stealing then why is he not the dream lead-off hitter? It comes down to hitting opportunites. It is known that the higher a player is in the lineup, the more plate appearances they will have. Imagine a scenario with the game on the line and two outs and the top of your lineup comes up. It would be much more favorable to have Cha hitting in that situation than Mertz. One last thing, he leads the team by a wide margin in strikeouts and strikeouts looking with 27 and 12 with the next highest being 15 and 8.

Inn - Inn is very similar to Mertz in that he is second in steals on the team with 21, has a similar OBP (.431), although Inn has a higher average and slugging percentage. Inn is a very control the zone hitter who only struck out 6 times in in 102 plate appearances (lowest on the team), walks 2.5 times more than he strikes out, and has the highest contact percentage with 92.1%. His problem is his 12.5% line drive rate and his high fly ball rate (45.83%) which is the lowest on the team and 3rd highest on the team respectively.

Hong - Hong did have the lowest amount of plate appearances in the dataset with 49 but did hit the 4th most doubles on the team with 5 in that short time span. He is one of only three players on the team with a slug over .400 (.415) but he also has the worst walk to strikeout ratio with .700. This strikeout rate is acceptable however due to his 2nd highest ISO (.122).

Fang - Fang had the second highest OBP on the team with (.561) but at the same time had a .250 average and a .271 slugging so the OBP was doing the heavy lifting in his great .832 OPS. He accomplishes this statline with 25 walks (3rd most) and only 12 hits (1 double). He was 2nd on the team with 15 strikeouts and 8 of them were looking. His BABIP was .364 suggesting that he was getting hits when he put the ball in play, it was just a matter of putting the ball in play that was difficult. Although, his 51.5% flyball rate is concerning especially for a player that was slugging under .300.

Bryson - Bryson had a very similar slashline to Inn with Bryson having the same amount of walks as Inn (15) and only 4 more strikouts (3rd lowest on team). One area of weakness for Bryson is his .167 average with runners in scoring position which is coupled with his 0 two-out RBI. Although Bryson did have a knack for seeing pitches as he saw the 3rd most pitches with 362, had the 3rd most plate appearances with 6 or more pitches with 19, and had the second most plate appearances where the batter saw 3 or more pitches after getting to two strikes with 14.

Blankenship - Blankenship is just outside of the top 3 in terms of slugging percentage with .394 and only struck out 7 times in 83 plate appearances. He also possesses the 2nd highest contact rate with 90.1% which is basically Inn with better slugging numbers and worse walk numbers. His 25% line drive rate is excellent and he is one of the more clutch hitter with a .343 average with RISP, 8 two-out RBI, and he also had the second most XBH on the team with 7.

Lie - Lie is an overall average player as his 261/393/275 slash line is close to the team's average slashline if we removed Nussbaum's numbers. He walks as much as he strikes out and his line drive rate is the second lowest on the team with 14.3%. He is above average at seeing pitches as he is 3rd with pitches seen per plate appearance at 4, and has 17 plate appearances with 6 or more pitches. He also does have 7 two-out RBI which would be 4th. His BABIP is .321 so the numbers suggest a bit of regression especially with the low line drive rate.

Robinson - Statistically the worst hitter on the team with the highest groundball rate at 48.7%. I couldn't find anything to suggest any luck factor or turnaround.

Nussbaum - One of the best three hitters on the team with the highest SLG (.662), OPS (1.145), wOBA (.480), RC (23.0), and ISO (.282). He walks more than he strikes out and has the most XBH and two-out RBI on the team with 13 and 14. His lower line drive rate is concerning at just 14.75%, however a player that slugs how he does should hit the ball in the air more often, which he does at a 54.1% fly ball rate.

8 Creating the Lineup

In the traditional sense of creating lineups, the typical lineup looks as follows:

1. Fastest Player
2. Contact Player/Good Bunter
3. Overall Best Hitter
4. Most Powerful Hitter
5. Good Hitter with lots of RBI
6. Next Best Hitter
7. Next Best Hitter
8. Next Best Hitter
9. Worst Hitter

The sabermetric way of creating a lineup sometimes uses elements of the traditional lineup but is more willing to sacrifice these traditional qualities of each spot in order to theoretically maximize runs. The core of this lineup would be designating your best 3 hitters on your team and recognize their strengths. The sabermetric looks as follows:

1. Highest OBP, one of best hitters
2. Best overall/balanced hitter, must have good OBP
3. Good overall batter but probably your 4th best batter
4. One of best hitters with the most power (SLG and ISO)
5. Good batter with power numbers
6. Next Best Hitter
7. Next Best Hitter
8. Next Best Hitter
9. Out of the 6-9 spots, whoever has the highest OBP (ability to steal is preferred)

In our lineup, the three best hitters are Cha, Nussbaum, and Hong. I will base the rest of the lineup around them and explain why each player is in each spot.

Sabermetric lineup, using all 10 players, not just 9.

1. Cha
2. Hong

3. Blankenship
4. Nussbaum
5. Inn
6. Bryson
7. Fang
8. Lie
9. Robinson
10. Mertz

9 Lineup Reasoning

Cha - One of the best hitters who has the highest OBP.

Hong - One of the best hitters but didn't have the highest OBP or SLG.

Blankenship - Aside from the best 3, had the second highest OPS (behind Fang) but had the highest SLG.

Nussbaum - One of the best hitters who had the best power numbers (SLG & ISO).

Inn - Tough decision on whether to put Inn or Bryson in this spot, decided it should be Inn because in similar plate appearances, he has the higher triple slash as well as wOBA, RC, and ISO.

Bryson - The next best player based on SLG.

Fang - Next best player, but has a high OBP making me consider putting him in the 10 hole, however only a 6.6 RC lead me to put Mertz in the 10 hole and Fang just wherever in the lineup that he falls.

Lie - Next best player based on wOBA, RC, SLG, ISO.

Robinson - Statistically the worst player.

Mertz - Not the worst player, but has the highest OBP to go along with 26 steals in 17 attempts (96.3% success rate). Great tablesetter for when the lineup flips over.

10 Conclusion

Without taking into account any physical attributes, I was able to use data to figure out the hitting profile of this team and construct what I believe to be the best lineup for manufactuing runs. Obviously, this does not guarentee more runs being scored, it just gives the team the best chance at scoring as many runs as possible on any given day.

The construction of the sabermetic lineup probably looks different from the one that is being deployed currently as this data was last updated in March of 2023 and there is new data being stored in a different GameChanger team. Barring some sort of drastic physical changes, do believe that this lineup would still hold up three months later and would love to construct another lineup using more recent stats from the spring of 2023.