LS88: Sports Analytics

Linear Weights & Efficiency

Linear Weights

Recall BA, OBP, and SLG

➔ Weight events by a value, and then add up (linear)

Linear Weights

Batter_ID	Lineup_Order	Inning	Outs	Event_Type	Start_Bases	End_Bases	Event_Runs	Run_Expectancy	Run_Expectancy_Next	RE24
seguj002	1	1	0	BB	None on	1st	0	0.529067	0.913562	0.384494
hanim001	2	1	0	K	1st	1st	0	0.913562	0.526186	-0.387376
canor001	3	1	1	Generic out	1st	1st	0	0.526186	0.231741	-0.294445
cruzn002	4	1	2	K	2nd	2nd	0	0.330982	0	-0.330982
escoy001	1	1	0	1B	None on	1st	0	0.529067	0.913562	0.384494
calhk001	2	1	0	2B	1st	2nd and 3rd	0	0.913562	2.0538	1.14024
troum001	3	1	0	Generic out	2nd and 3rd	2nd	1	2.0538	0.718396	-0.335402
pujoa001	4	1	1	Generic out	2nd	2nd	0	0.718396	0.330982	-0.387414
cronc002	5	1	2	Generic out	2nd	2nd	0	0.330982	0	-0.330982
seagk001	5	2	0	Generic out	None on	None on	0	0.529067	0.279063	-0.250004

From Retrosheet, we have RE24 for every plate appearance

Linear Weights

For each type of event, average the observed RE24 values

This is *Linear Weights*: average run production values by event

- ➔ We can use them to directly weight events
- ➔ Weight values give contribution in *runs*, not some other unit

Later, we'll use the weights to build wOBA

Linear Weights Interpretation

- ➔ Randomly pick a single that occurred during the season
- ➔ What do we expect this single to have produced in runs?

- ➔ Let’s use the single as an example

  Single with no one on: just adds a runner at first. That’s an easy value boost to compute.
  Single with runner on first: sometimes leads to 1st and 2nd, sometimes 1st and 3rd
  Single with the bases loaded: sometimes 1 run, sometimes 2, maybe 3?? Maybe runner out at home?

  - ◆ 1st and 3rd is better than 1st and 2nd
  - ◆ Singles with bases loaded are very high value

Linear Weights Interpretation

$$\text{Run Value for Event } E = \sum_{\text{All Outcomes}} \{\text{Prob. of Outcome } O \text{ from Event } E\} \times \{\text{RE24 of Outcome } O\}$$

The average balances frequency of situation/baserunning outcomes against observed run production values

- 1st and 3rd is better but maybe not as likely as 1st and 2nd
- Bases loaded is a rarer situation, so the high value is tempered by rarity (a runner thrown out at home is even rarer)

Interpreting the Values

Event_Type	Generic out	K	FC	IBB	BB	HBP	Interference	1B	RBOE	2B	3B	HR
RE24	-0.284	-0.279	-0.233	0.182	0.314	0.335	0.399	0.449	0.491	0.779	1.056	1.38

All values are net runs produced above average

- ➔ A home run is worth +1.38. Why?

  Accounts for frequency of runners on base.

- ➔ Intentional walk worth less than a regular walk

  Typically done when first base is open. Can lead to more force out opportunities. Also, who's getting IBBed?

- ➔ HBP is worth slightly more than a walk

wRAA

wRAA: Weighted Runs Above Average

	Double	Generic out	Hit by pitch	Home run	Intentional walk	PA	Single	Triple	Walk
Barry Bonds (2001)	32	320	9	73	35	664	49	2	177

$$122.8 \text{ Runs Above Average} = 0.778 \cdot 32 - 0.283 \cdot 320 + 0.336 \cdot 9 + 1.379 \cdot 73 + 0.17 \cdot 35 + 0.461 \cdot 49 + 1.081 \cdot 2 + 0.306 \cdot 177$$

LWTS and wRAA are also known as Batting Runs by Pete Palmer

Pete Palmer fixes the positive weights and lets the out value fluctuate to account for variable run scoring environments

wRAA

If you multiply the weights by frequencies of the events and sum:

Result is (near) 0

Why?

- ➔ When building RE24, we subtract an expectation: we're computing *changes* in expectation
  Also known as the marginal impact
- ➔ The linear weights are telling you about *net runs produced above/below average.*
  An event with weight 0 would maintain average scoring rate
- ➔ Weight tells you the incremental expected run value of an event

wRAA tells a similar story to RE24 but we remove the situational dependence

wOBA

We want to rate a player on events *relative to outs*

BA, OBP, and SLG set the value of an out at 0.

We do the same for wOBA. How?

➔ Compute the difference of the value of an event with the value of an out

$$\text{Relative Value} = \text{LWTS Event Value} - \text{LWTS Out Value}$$

wOBA

One last thing to do for wOBA: the wOBA scale

wOBA scale is used to alter the presentation to make it similar scale to OBP

- ➔ For most people, quoting runs relative to outs per PA is tough to understand
- ➔ We do however have a feel for good, average, and bad OBP values

$$\text{wOBA Value} = (\text{LWTS Event Value} - \text{LWTS Out Value}) \times \text{wOBA Scale}$$

Season	wOBA	wOBAScale	wBB	wHBP	w1B	w2B	w3B	wHR
2017	.321	1.185	.693	.723	.877	1.232	1.552	1.980
2016	.318	1.212	.691	.721	.878	1.242	1.569	2.015
2015	.313	1.251	.687	.718	.881	1.256	1.594	2.065

wOBA

We have the wOBA weights, we can build the wOBA value

This is similar to how we build BA, OBP, or SLG

Event_Type	Walk	Hit by pitch	Single	Double	Triple	Home run
RE24	0.696	0.732	0.879	1.254	1.613	1.965

wOBA weights for 2001 (wOBA scale 1.182)

$$wOBA = (0.696 \cdot BB + 0.732 \cdot HBP + 0.879 \cdot 1B + \\ 1.254 \cdot 2B + 1.613 \cdot 3B + 1.965 \cdot HR)/PA$$

Linear Weights and wOBA

Linear Weights (LWTS)

Incremental change in expected runs by an event. Events like hits are a plus. Outs are naturally minus.

wOBA Weights

Incremental value of an event *relative to an out*.

Is a home run 3x more valuable than a single or 2.2x more valuable?

LWTS says 3, wOBA says 2.2. Which is right?

LWTS. Why?

Linear Weights and wOBA

wOBA weights quantify the tradeoff between the events and an out.

Ex: Convert 2.2 PA from an out to a single to have the same impact as 1 PA to a home run

Better and more usual to think in terms of LWTS values.

Another way to think about them/how to use them:

- Adding a walk to a batter’s line (and therefore also adding a PA)
  - *Use the LWTS value: ~0.32 runs*
- Exchanging a walk for an out in a batter’s line (keeping total PA the same)
  - *Use the unscaled wOBA value: ~0.62 runs*

Classical Stats, LWTS, and wOBA

Let’s revisit the classic stats one last time...

Let’s also recall OPS:

$$OPS = OBP + SLG$$

Classical Stats, LWTS, and wOBA

A little bit of algebra applied to OPS...

$$OPS = OBP + SLG$$

$$\Rightarrow$$

$$OPS = \frac{1}{PA}\left(2 \cdot \mathit{1B} + 3 \cdot \mathit{2B} + 4 \cdot \mathit{3B} + 5 \cdot HR + 1 \cdot BB\right) + \text{Other Stuff}$$

$$\Rightarrow$$

$$OPS = \frac{4}{PA}\left(0.50 \cdot \mathit{1B} + 0.75 \cdot \mathit{2B} + 1.00 \cdot \mathit{3B} + 1.25 \cdot HR + 0.25 \cdot BB\right) + \text{Other Stuff}$$

OPS has a linear component with weights that look kind of familiar!

Classical Stats, LWTS, and wOBA

Weights comparison

	$1B$	$2B$	$3B$	HR	BB	HBP	O
BA	1	1	1	1			
OBP	1	1	1	1	1	1	
SLG	1	2	3	4			
OPS	2	3	4	6	1	1	
$LWTS$	0.48	0.78	1.07	1.40	0.32	0.32	-0.3
$wOBA$	0.879	1.254	1.613	1.965	0.696	0.732	

Classical Stats, LWTS, and wOBA

Weights comparison with a bit of rescaling

	$1B$	$2B$	$3B$	HR	BB	HBP	O
BA	1	1	1	1			
OBP	1	1	1	1	1	1	
SLG	0.333	0.667	1.000	1.333			
OPS	0.500	0.750	1.000	1.250	0.250	0.250	
$LWTS$	0.461	0.778	1.081	1.379	0.306	0.336	-0.283
$wOBA$	0.879	1.254	1.613	1.965	0.696	0.732	

- ➔ SLG, OPS, and LWTS are all close, but only LWTS has a value for an out
- ➔ Add out value and scale by 1.182 to SLG/OPS weights and you get close to wOBA

Efficiency

Resources

A fact of life, and sports, is *limited resources*

In baseball, we only get 3 outs per inning and 27 in a 9 inning game

Sabermetrics has changed how teams think about using the limited resources

- ➔ No longer willingly giving up outs for baserunner advancement
- ➔ In important, aka high leverage situations, teams deploy their best pitchers
  The outs are more valuable given the context
- ➔ Starters are still valuable because roster spots are limited
  A team can't use 20 pitchers to each go 1 inning

Resources

To be a great hitter, one must balance

- ➔ Not wasting PA by swinging at bad pitches and thus taking a walk
- ➔ Hitting safely with a high BA
- ➔ Hitting for power to do more damage

A hitter needs to produce big upside (power), but do it efficiently by downsides (outs)

Inefficient Batting

Rougned Odor in 2017 produced power, but not efficiently

Year	Age	Tm	Lg	G	PA	AB	R	H	2B	3B	HR	RBI	SB	CS	BB	SO	BA	OBP	SLG	OPS	OPS+	TB	GDP	HBP	SH	SF	IBB	Pos	Awards
2014	20	TEX	AL	114	417	386	39	100	14	7	9	48	4	7	17	71	.259	.297	.402	.698	93	155	7	5	6	3	1	*4/D	
2015	21	TEX	AL	120	470	426	54	111	21	9	16	61	6	7	23	79	.261	.316	.465	.781	107	198	3	14	2	5	2	*4	
2016	22	TEX	AL	150	632	605	89	164	33	4	33	88	14	7	19	135	.271	.296	.502	.798	105	304	6	4	0	4	0	*4/D	
2017	23	TEX	AL	***162***	651	607	79	124	21	3	30	75	15	6	32	162	.204	.252	.397	.649	63	241	13	8	0	4	5	*4/D	
2018	24	TEX	AL	129	535	474	76	120	23	2	18	63	12	**12**	43	127	.253	.326	.424	.751	95	201	5	11	2	5	2	*4/D	
5 Yrs				**675**	**2705**	**2498**	**337**	**619**	**112**	**25**	**106**	**335**	**51**	**39**	**134**	**574**	**.248**	**.295**	**.440**	**.735**	**91**	**1099**	**34**	**42**	**10**	**21**	**10**		
162 Game A...			**.**	**162**	**649**	**600**	**81**	**149**	**27**	**6**	**25**	**80**	**12**	**9**	**32**	**138**	**.248**	**.295**	**.440**	**.735**	**91**	**264**	**8**	**10**	**2**	**5**	**2**		

Resources in Basketball

Unlike baseball, a basketball game (and most other sports) has a time limit

Hence, time is a limited resource

We can also think of possessions as a limited resource

- ➔ Without off. rebounds and turnovers, each team would have the same number of possessions
- ➔ Even if the total is variable depending on pace of play, a team doesn’t get more than the other

Resources in Basketball

Efficient usage of time/possessions is a “recent” revolution

- Research lagged behind baseball
- Many teams/analysts/fans slow to adopt

It doesn’t help that Kobe in his prime, and in the conversation for best player each year, was also (kind of) the model for inefficiency

- Russell Westbrook has taken the torch from Kobe

Inefficient Basketball

The quintessential Kobe game (April 13, 2016)

Lakers

STARTERS	MIN	FG	3PT	FT	OREB	DREB	REB	AST	STL	BLK	TO	PF	+/-	PTS
J. Randle PF	33	1-4	0-0	0-0	0	9	9	1	1	1	1	2	+3	2
K. Bryant SF	42	22-50	6-21	10-12	0	4	4	4	1	1	2	1	+7	60
R. Hibbert C	21	2-3	0-0	0-0	3	3	6	2	0	0	1	2	+3	4
J. Clarkson PG	34	6-10	0-1	0-0	2	5	7	1	1	0	1	1	+9	12
D. Russell PG	36	4-10	0-3	1-1	0	4	4	5	2	0	2	3	+1	9

Quantifying Efficient Basketball

So how do we quantify efficient basketball?

If time or possessions are the limited resource, then we need to compute our performance per unit in that resource

Points per Game it out.

Points per Possession is in!

- For players, we could consider Points per Minute Played
- Possession is the better choice: pace of play doesn't get manipulated

Quantifying Efficient Basketball

Points per Possession is a very useful way to measure efficiency

- > 1 is good

We can also use Off. Rating: Points per 100 Possessions

- Just multiply PPP by 100

Counting Possessions

Come on, possessions are easy to count in basketball

Soccer is the hard sport!

Counting possessions is easy if you have play-by-play data

Unfortunately, we don't have play-by-play data going back to olden times (90s)

Two choices:

- ➔ Restrict to the recent era with PxP data: fine if you narrow your focus
- ➔ Estimate the number of possessions

Estimating Possessions

From Basketball Reference

$$\begin{aligned}\text{Off. Poss.} = {} & \text{TmFGA} + \\ & 0.4 \times \text{TmFTA} - \\ & 1.07 \times (\text{TmORB}/(\text{TmORB} + \text{OppDRB})) \times (\text{TmFGA} - \text{TmFG}) + \\ & \text{TmTOV}\end{aligned}$$

Similar for Def. Poss.

$$\text{Possessions} = \frac{\text{Off. Poss.} + \text{Def. Poss.}}{2}$$

Measures for Efficiency

We can quantify efficiency with Off. Rating

A problem though: it’s global. Every element of offensive play affects it

Let’s look at ways to isolate a key component of efficient play:

Shooting Percentage

Field Goal Percentage

The most obvious and classical to consider is Field Goal%

$$\text{FG }\% = \frac{\text{Field Goals}}{\text{Field Goals Attempts}}$$

We’re going to uncover the following analogy

FG% : Basketball :: Batting Average : Baseball

3 > 2: Effective Field Goal Percentage*

Aka eFG%, it's considered an "advanced stat"

$$\text{Effective FG }\% = \frac{\text{Field Goals} + \frac{1}{2}\cdot\text{3-pt Field Goals}}{\text{Field Goals Attempts}}$$

We all know 3 > 2, so give the credit!

Another analogy:

eFG% : Basketball :: Slugging Pct : Baseball

*Good Janet says, "not a percentage"

Charity Stripe: True Shooting Percentage*

We care about *using* possessions to score, so why are ignore Free Throws?

$$\text{True Shooting \%} = \frac{\text{Total Points}}{2(\text{Field Goals Attempts} + .44 \cdot \text{Free Throw Attempts})}$$

Divide by 2 to make it close to the others

Without the 2, it measures Points per Possession Used to Shoot

Another analogy

TS% : Basketball :: wOBA :: Baseball

*Good Janet says, "also not a percentage"

Shooting Percentage Summary

“Totals” over “opportunities”

FG%: Makes per Attempts from the Floor

eFG%: Points per Field Goal Attempt

Efficiency from the floor (ignores FTs!)

TS%: Total points per possessions used to shoot

- ➔ Turnovers can count as a possession used. That’s ignored for now
- ➔ Where does the .44 value in front of FTA come from?

  FTs typically come in pairs so .5 x FTs is approximately the number of possessions used that ended in 2 FT shots. .44 accounts for flagrants, technicals, And-1s, 3pt fouls, etc

Top 20 Performance

Field Goal Pct	
1. Clint Capela • HOU	.656
2. DeAndre Jordan • LAC	.653
3. Steven Adams • OKC	.634
4. Enes Kanter • NYK	.607
5. John Collins • ATL	.591
6. John Henson • MIL	.589
7. Taj Gibson • MIN	.574
8. Jonas Valanciunas • TOR	.569
9. Hassan Whiteside • MIA	.554
10. Andre Drummond • DET	.552
11. Julius Randle • LAL	.551
12. Derrick Favors • UTA	.550
13. Dwight Howard • CHO	.547
14. Giannis Antetokounmpo • MIL	.544
15. LeBron James • CLE	.542
Karl-Anthony Towns • MIN	.542
17. Anthony Davis • NOP	.542
18. Domantas Sabonis • IND	.538
19. Marcin Gortat • WAS	.537
20. Ben Simmons • PHI	.526

Effective Field Goal Pct	
1. Clint Capela • HOU	.656
2. DeAndre Jordan • LAC	.653
3. Steven Adams • OKC	.634
4. Joe Ingles • UTA	.615
5. Stephen Curry • GSW	.615
6. Enes Kanter • NYK	.607
7. E'Twaun Moore • NOP	.604
8. John Collins • ATL	.599
9. Karl-Anthony Towns • MIN	.595
10. Klay Thompson • GSW	.593
11. Joe Harris • BRK	.592
12. John Henson • MIL	.589
13. Jonas Valanciunas • TOR	.589
14. LeBron James • CLE	.588
15. Trey Lyles • DEN	.587
16. Kevin Durant • GSW	.585
17. Taj Gibson • MIN	.581
18. Al Horford • BOS	.578
19. Gary Harris • DEN	.578
20. Kelly Olynyk • MIA	.573

True Shooting Pct	
1. Stephen Curry • GSW	.673
2. Clint Capela • HOU	.661
3. DeAndre Jordan • LAC	.656
4. Mike Scott • WAS	.645
5. Enes Kanter • NYK	.641
6. Steven Adams • OKC	.640
7. Darius Miller • NOP	.639
Karl-Anthony Towns • MIN	.639
9. Kyle Korver • CLE	.637
10. Jonas Valanciunas • TOR	.635
11. Kevin Durant • GSW	.634
12. John Collins • ATL	.632
13. Marvin Williams • CHO	.624
14. Joe Ingles • UTA	.623
15. LeBron James • CLE	.621
16. Montrezl Harrell • LAC	.621
17. Anthony Davis • NOP	.620
18. James Harden • HOU	.619
19. Nikola Mirotic • TOT	.614
20. Giannis Antetokounmpo • MIL	.613
Kevin Love • CLE	.613

Top 20 Performance

TS% can change our view of a player's performance a lot

- FG% loves big men: short shots
  DeAndre Jordan 2016-17: #2 all time in FG% and eFG%, #18 all time in TS%
- eFG% boosts up 3pt shooters
- James Harden finally in the top 20 for TS%
- Makes sense: an attacking player getting a lot of FTs
  - Inherently shoots tough shots so FG% is going to be low
  - But if the player earns fouls and scores from the line, that needs to count
    That's a good use of a possession! (this is like counting a walk)

Our Analogies

Recall our analogies

FG% : Basketball :: Batting Average : Baseball

eFG% : Basketball :: Slugging Pct : Baseball

TS% : Basketball :: (sort of) wOBA :: Baseball

Weighting differently and accounting for an important component of scoring directly translate from baseball

The analogies also indicate performance as a metric

Shooting Percentage Demo