Fantasy Baseball Draft Tool

Documentation and Methodology

Claude + Kevin

February 2026

Outline

What Does This Tool Do?

Goal: Rank players by their marginal contribution to winning H2H categories.

Key Components:

1. `create_league_stats.py` – Generates hitter CSVs with z-scores
2. `create_pitching_stats.py` – Generates pitcher CSVs
3. `normalize_pa.py` – Aligns PA between projection systems
4. `draft_tool.html` – Interactive draft UI with real-time valuations

League Format: 14-category H2H weekly

- Hitting: R, HR, RBI, SB, SO, TB, OBP
- Pitching: W, SV, K, HLD, ERA, WHIP, QS

The Core Idea: Marginal Win Probability

In H2H, you win a category if your weekly total beats your opponent's.

Assuming team totals are normally distributed:

$$P(\text{win category}) = \Phi\left(\frac{\mu_{\text{my team}} - \mu_{\text{opponent}}}{\sigma \cdot \sqrt{2}}\right)$$

where:

- Φ = standard normal CDF
- σ = weekly standard deviation for that category (from 2024 data)
- The $\sqrt{2}$ comes from $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) = 2\sigma^2$

A player's value = how much they move $P(\text{win})$ vs. replacement.

Weekly Standard Deviations (2024 League Data)

Hitting Categories:

Category	SD	Mean	CV	1/SD
R	6.03	28.96	21%	0.17
HR	2.93	8.02	37%	0.34
RBI	6.72	27.86	24%	0.15
SB	2.57	4.74	**54%**	**0.39**
SO	7.45	50.11	15%	0.13
TB	15.94	88.87	18%	0.06
OBP	0.04	0.320	13%	25.0

Two metrics:

- **1/SD**: Marginal value per unit (used for player valuation)
- **CV = SD/Mean**: How luck-dependent the category is

SD vs Coefficient of Variation

Why does absolute SD matter for marginal value?
The win probability formula is:

$$P(\text{win}) = \Phi\left(\frac{\mu_{me} - \mu_{opp}}{\sigma\sqrt{2}}\right)$$

Adding $+1$ unit shifts numerator by $+1$. Denominator is $\sigma\sqrt{2}$.
The mean doesn't appear – only absolute SD matters for marginal impact.

But CV tells us something different:

- SB has CV $= 54\%$ $\rightarrow$ outcomes highly variable relative to baseline
- SO has CV $= 15\%$ $\rightarrow$ outcomes more predictable

High CV means the category is "noisy," but the marginal value of $+1$ unit is still determined by $1/\text{SD}$.

Pitching Standard Deviations

Category	SD	Mean	CV	1/SD
L	1.83	3.08	59%	0.55
SV	1.54	2.27	**68%**	**0.65**
K	11.79	50.90	23%	0.08
HLD	1.64	2.30	**71%**	0.61
ERA	1.31	3.79	35%	0.76
WHIP	0.21	1.20	17%	4.76
QS	1.40	3.25	43%	0.71

SV and HLD have **both** high leverage (high 1/SD) **and** high noise (high CV).
These categories are volatile but also where edges compound.

How Z-Scores Are Calculated in CSVs

For **counting stats** (R, HR, RBI, TB, SB):

$$z = \frac{\text{Season Total}/25}{\text{Weekly SD}}$$

For **strikeouts** (lower is better):

$$z_{SO} = -\frac{\text{Season SO}/25}{\text{SD}_{SO}}$$

For **OBP** (rate stat):

$$z_{OBP} = \frac{(\text{OBP} - 0.320)}{9 \times \text{SD}_{OBP}}$$

Why divide by 9? OBP is a rate stat. A roster has 9 hitters, so one player's OBP contributes $\frac{1}{9}$ of team OBP. This scales OBP's contribution appropriately.

Total Z-Score

$$z_{\text{total}} = z_R + z_{HR} + z_{RBI} + z_{SB} + z_{TB} + z_{SO} + z_{OBP}$$

This is a simplified ranking metric used to sort players in the CSV.

Important distinction:

- **CSV z-scores**: Quick approximation for ranking
- **Draft tool marginal value**: Exact calculation considering current roster

Both should *mostly* agree on rankings, but the draft tool is more precise because it accounts for your specific team composition.

What Is Replacement Level?

Definition: The production available from freely available players (waiver wire).

Methodology:

1. Rank all hitters by z_{total} using Depth Charts projections
2. Take players ranked 155–175 (just beyond 16 teams $\times$ 9 hitters $=$ 144)
3. Average their per-PA rates

The cohort (21 players):

Spencer Steer, Miguel Andujar, Ezequiel Tovar, Jonathan Aranda, Addison Barger, Nathan Lukes, Kyle Manzardo, Colt Keith, Josh Lowe, Romy Gonzalez, Samuel Basallo, Francisco Alvarez, Lars Nootbaar, Joey Ortiz, Tyler O'Neill, Ryan O'Hearn, Victor Robles, Jake Fraley, JJ Bleday, Munetaka Murakami, Chase Meidroth

Replacement Level Per-PA Rates

Stat	Per-PA Rate	Full Season (600 PA)	Weekly
R	0.1212	73	2.92
HR	0.0330	20	0.80
RBI	0.1208	72	2.88
SO	0.2226	134	5.36
TB	0.3725	224	8.96
SB	0.0152	9	0.36
OBP	0.320	–	–

OBP is hardcoded to 0.320 (league average) because the cohort's actual OBP (0.324) exceeded league average. This would make replacement players OBP-positive, which is conceptually wrong—replacement should be neutral, not additive.

PA Supplementation

Problem: Low-PA players look bad in counting stats.

Solution: Supplement everyone to 600 PA with replacement-level production.

For a player with 450 PA:

$$\begin{aligned} \text{Gap PA} &= 600 - 450 = 150 \\ \text{Runs} &= \text{Projected Runs} + 150 \times 0.1212 \\ \text{OBP} &= \frac{450 \times \text{proj OBP} + 150 \times 0.320}{600} \end{aligned}$$

Interpretation: "What would this player produce if they played full-time, with a replacement-level player filling in the rest?"

Setup: A Team of Replacement Players

Consider a team with 9 replacement-level hitters.

Each replacement hitter's weekly production:

Stat	Per Hitter/Week
R	2.92
HR	0.80
RBI	2.88
SO	5.36
TB	8.96
SB	0.36
OBP	0.320

Team totals $= 9 \times$ individual $=$ (26.3 R, 7.2 HR, 25.9 RBI, 48.2 SO, 80.6 TB, 3.2 SB, .320 OBP)

Replacement Team vs League Average Opponent

League averages (expected opponent production):

Category	My Team	Opponent	Diff
R	26.28	28.96	-2.68
HR	7.20	8.02	-0.82
RBI	25.92	27.86	-1.94
SO	48.24	50.11	-1.87 (good!)
TB	80.64	88.87	-8.23
SB	3.24	4.74	-1.50
OBP	0.320	0.320	0

A replacement-level team is *below average* in most categories.
The exception: SO is below average (good, since lower is better).

Win Probabilities: Replacement Team

Using $P(\text{win}) = \Phi\left(\frac{\text{diff}}{\sigma\sqrt{2}}\right)$:

Category	Diff	SD	z	P(win)
R	−2.68	6.03	−0.31	37.7%
HR	−0.82	2.93	−0.20	42.2%
RBI	−1.94	6.72	−0.20	41.9%
SO	+1.87	7.45	+0.18	57.0%
TB	−8.23	15.94	−0.37	35.8%
SB	−1.50	2.57	−0.41	34.0%
OBP	0	0.04	0	50.0%

Expected category wins: $0.377 + 0.422 + \ldots = 2.99$ out of 7.

Enter José Ramírez

Ramírez's projected season line:
679 PA, 98 R, 30 HR, 94 RBI, 78 SO, 300 TB, 34 SB, .348 OBP

Ramírez's weekly production:

Stat	Ramírez	Replacement	Diff
R/wk	3.92	2.92	+1.00
HR/wk	1.20	0.80	+0.40
RBI/wk	3.76	2.88	+0.88
SO/wk	3.12	5.36	−2.24 (good!)
TB/wk	12.00	8.96	+3.04
SB/wk	1.36	0.36	+1.00
OBP	0.348	0.320	+0.028

Team After Adding Ramírez

Replace one replacement hitter with Ramírez:

Category	Before	After	Change
R	26.28	27.28	$+1.00$
HR	7.20	7.60	$+0.40$
RBI	25.92	26.80	$+0.88$
SO	48.24	46.00	-2.24
TB	80.64	83.68	$+3.04$
SB	3.24	4.24	$+1.00$
OBP	0.320	0.323	$+0.003$

Note: OBP only shifts by $\frac{1}{9}$ of Ramírez's OBP advantage because it's a rate stat averaged across 9 hitters.

New Win Probabilities

Cat	Old Diff	New Diff	Old P	New P	ΔP
R	-2.68	-1.68	37.7%	42.2%	$+4.5\%$
HR	-0.82	-0.42	42.2%	46.0%	$+3.8\%$
RBI	-1.94	-1.06	41.9%	45.6%	$+3.6\%$
SO	$+1.87$	$+4.11$	57.0%	65.2%	$+8.1\%$
TB	-8.23	-5.19	35.8%	40.9%	$+5.1\%$
SB	-1.50	-0.50	34.0%	44.5%	$+10.5\%$
OBP	0	$+0.003$	50.0%	52.2%	$+2.2\%$

Ramírez's marginal value:

$$\sum \Delta P = 4.5 + 3.8 + 3.6 + 8.1 + 5.1 + 10.5 + 2.2 = \mathbf{37.8}\%$$

Ramírez adds 0.378 expected category wins per week vs replacement.

Why SB and SO Dominate

Ramírez's biggest impacts:

- **SB: +10.5%** from just +1.00 SB/week
- **SO: +8.1%** from −2.24 SO/week

Compare to HR: +3.8% from +0.40 HR/week.

The math:

$$\text{SB impact} : +1.00/(2.57 \times \sqrt{2}) = +0.275 \text{ z-shift}$$
$$\text{HR impact} : +0.40/(2.93 \times \sqrt{2}) = +0.097 \text{ z-shift}$$

The steal is worth $0.275/0.097 = 2.8\times$ more in z-shift than the HR, despite both being "+1 unit" contributions (after scaling).

Key insight: Low absolute SD $\rightarrow$ high marginal leverage.

Visualizing the Distribution Shift

Before Ramírez (SB category):

- My team mean: 3.24 SB/week
- Opponent mean: 4.74 SB/week
- Difference: -1.50, so $z = -0.41$, $P(\text{win}) = 34\%$

After Ramírez:

- My team mean: 4.24 SB/week
- Opponent mean: 4.74 SB/week
- Difference: -0.50, so $z = -0.14$, $P(\text{win}) = 44.5\%$

The entire distribution of "my SB − opponent SB" shifts right by 1.00 steals. Because SD is tight (2.57), this 1-steal shift moves us **10.5 percentage points** in win probability.

RP Replacement Level (Per Slot, Weekly)

Stat	Replacement RP
IP/week	2.48
L/week	0.118
SV/week	0.121
HLD/week	0.848
K/week	2.69
ER/week	0.96
WH/week	2.85

Key observation: Replacement RPs get **holds, not saves**.
Elite closers get saves but sacrifice holds. This creates an explicit tradeoff that the model captures.

Why ERA/WHIP Don't Matter Much for RPs

ERA/WHIP are innings-weighted ratio stats.

Typical team: 40 IP/week total

- 5 SP $\times$ 6.5 IP $=$ 32.5 IP
- 3 RP $\times$ 2.5 IP $=$ 7.5 IP

A reliever contributes $\approx$**6–7%** of team innings.

Even if an elite RP has much better ERA than replacement (e.g., 3.06 vs 3.50), the team ERA only improves by $\approx$0.02 points.
With ERA SD $= 1.31$, that's $0.02/1.31 = 0.015$ SDs $\rightarrow$ $<$**1% win probability**.

Conclusion: RPs earn value through saves/holds/K, not ERA/WHIP.

The Problem: Projection Systems Disagree on PT

Different projection systems predict different playing time:

Player	The Bat PA	Depth Charts PA
Player A	550	620
Player B	480	510

Problem: We want to compare *skill*, not playing time estimates.

Solution: Normalize all systems to use Depth Charts PA.

PA Normalization Process

`normalize_pa.py`:

1. Read Depth Charts PA for each player
2. For each player in The Bat/BatX:
   - Scale = DC_PA / TheBat_PA
   - Multiply all counting stats by Scale
   - Keep rate stats (OBP, K%) unchanged
3. Output normalized CSV

Keeping rate stats unchanged is reasonable—the rate estimates reflect the projection system's view of player skill, independent of playing time.

Three Projection Systems Available

1. **Depth Charts** (default) – Composite of multiple systems
2. **The Bat** – Tom Tango's projection system
3. **The BatX** – Extended/experimental version

All three use:

- Same PA (normalized to Depth Charts)
- Same replacement level rates
- Same weekly SDs

Only difference: Rate stat projections (HR/PA, SB/PA, K%, OBP, etc.)

Open Questions for Review

1. **[?] Choice of ranks 155-175** – Why not 145-165 or 160-180? The exact cutoff affects replacement level.
2. **[?] SP replacement level** – I didn't examine this closely. Need to verify the methodology matches hitter replacement.
3. **[?] Variance equality assumption** – The $\sqrt{2}$ assumes both teams draw from distributions with the same variance. In practice, team quality varies, but this is a reasonable simplification.

Feature Timeline

1. **Base draft tool** – UI, player data, basic rankings
2. **Marginal win probability** – Exact $\Phi()$ calculation
3. **PA supplementation to 600 PA floor** – Fill with replacement
4. **Projection toggle** – The Bat vs Depth Charts
5. **PA normalization** – All systems use DC playing time
6. **300 PA minimum filter** – Exclude low-PA players
7. **Replacement recalibration** – Ranks 155-175 cohort
8. **OBP cap at league average** – Prevent OBP-positive replacement
9. **The BatX projection system** – Third toggle option

Key Code Files

Python scripts:

- `create_league_stats.py` – Hitter z-scores, replacement supplementation
- `create_pitching_stats.py` – Pitcher processing
- `normalize_pa.py` – Align PA across systems

Main application:

- `draft_tool.html` – All-in-one HTML/CSS/JS app
  - Lines 482-506: SD and replacement constants
  - Lines 580-599: `normalCDF` and `winProbability`
  - Lines 604+: Roster projection calculations

Data files:

- `fantasy_hitters_dc_2026.csv`, `fantasy_hitters_thebat_2026.csv`, `fantasy_hitters_batx_2026.csv`

Summary

What the tool does:

- Calculates marginal win probability for each player vs replacement
- Uses weekly SDs from 2024 league data to weight categories
- Normalizes PA across projection systems for fair comparison
- Supplements low-PA players to 600 PA baseline

Key insight: Categories with low absolute SD (SB, SV, HLD) have outsized marginal impact. A single steal shifts win probability more than a single HR because $SD_{SB} < SD_{HR}$.