Latest News: **OOTP Update 21.6.83** - **OOTP 21 Available** - **FHM 7 Available**

FRANCHISE HOCKEY MANAGER 7 - AVAILABLE NOW!

User Name [User Name]   ☐ Remember Me?
Password [   ] [Log in]

Register   Blogs ▽   FAQ   Calendar   Today's Posts   Search

Perfect Team 22 Perfect Team 22 - The online revolution! Battle tens of thousands of PT managers from all over the world and become a legend.

[POST REPLY]

🔽 VIEW FIRST UNREAD   Thread Tools ▽

🟧 03-15-2021, 11:52 PM   #1

scipper
Bat Boy

Join Date: Nov 2018
Posts: 19

OOTP Statistical Analysis Guide

Intro

This is a guide to OOTP perfect team (PT) statistical analysis. For the past 2 years I've worked to come up with some systems for analyzing PT players, and I will share them here. Succeeding at PT involves many different aspects. I will talk about only 2 of them here: 1. Creating a good database of every player and 2. Projecting how players will perform. This isn't a perfect guide and there are others who have better systems, but this still contains some wisdom I've gathered over the past year. I spent a lot of time building this system, but the work takes too much time for me, so I won't be as active in OOTP 22. I wanted to share this to allow people to reproduce or build on what I've done. This guide will reproduce something like my system for yourself. It will be technical. For non-technical players, I will include recaps and learnings so that you can learn, too. This game contains a lot of information, not everything can be done alone. I encourage forming

small communities and sharing information.

For those interested in the technical parts, I'll have an example repo up written in python on github. See it
[here](#).

Last edited by scipper; 03-16-2021 at 12:15 AM. Reason: adding link to repo

QUOTE

03-16-2021, 12:01 AM   #2

scipper
Bat Boy

Join Date: Nov 2018
Posts: 19

Spreadsheet Database

These next few articles describe how to:
1. Pull from the PT card list (https://www.ootpdevelopments.com/per...all-card-list/)
2. Extract a reasonable spreadsheet.

We will be adding on to the publicly available data with our own custom metrics.

Player Attributes

TL;DR Basics of sheet generation and some information we want to show. Nothing super deep yet.

Almost every piece of information on a PT card gives useful information about a player. Let's talk about what
we care about in a sheet. This is a groundwork section for wrangling your spreadsheets.

Aside: How to pull info from the database
1. Go to https://www.ootpdevelopments.com/per...all-card-list/
2. Right click and hit "view source"
3. Search on that page for the line "var cards"
4. Highlight that whole line and copy it into a "Cards.txt" file. Remove everything but the "[]" and everything in
between (make sure to get the ";" at the end of the line)

Since baseball is a 2-phase game (offense and defense) there's a split between batters and pitchers. This
guide will describe how to create a generated spreadsheet.

For batters the ratings we care about are the main 5 (or 6) ratings for an at bat: 1. Gap Power (GAP) 2.
Home Run Power (POW) 3. Taking walks (EYE) 4. Avoiding strikeouts (AvK) 5. Batting average on balls in
play (BABIP). Note that there is no Contact (CON) rating. CON is calculated based on a formula of POW,
AvK, and BABIP. For analysis purposes the base ratings are better. People with a trained eye will generally
be able to tell a batter's strengths through a CON rating with no BABIP. Each of these ratings matters vL/vR.

We will discuss how to derive BABIP in a different article. Additionally, batting arm (right, left, switch) and batting profile (groundball/flyball tendency) make a difference.

Important for defense are the ratings for each position, including height (in cm) for 1B. Throwing arm will drop some players from being eligible for different positions.

Finally there are the baserunning ratings - speed, stealing, and baserunning. Speed and stealing don't work quite like expected, see baserunning analysis section.

For pitchers, we care about the big 3: Stuff (STU), Movement (MOV), and Control (CON). The throwing hand (lefty or righty) makes a large difference for vL/vR stats. Also relevant here is the groundball tendency of pitchers. These ratings affect what a pitcher gives up in an at bat. All pitchers have different ratings for stuff as a starter vs. stuff as a reliever. This is useful for converting starters to relievers.

For "defense" the ratings are the pitcher defense rating and hold rating. These affect steals upon the pitcher and some stuff around extra bases.

To see how we might go from the cards info from the database to a spreadsheet, see this branch.

League Pulls

TL;DR to be able to perform analysis, we'll need to pull a lot of information. Most stats OOTP makes available in a league download. We'll do this for the overall info, the versus left info, and the versus right info.

When gathering information about a league to analyze we want basically every stat that OOTP offers. The only ones we don't take should be the ones that we can calculate from others - e.g. if we know # of hits, # of home runs, doubles, and triples then we can calculate total bases. This is a lot of information to pull so you'll want to save this custom view.

The list of things to include is in the album.

Conduct pulls after the regular season is over. This is usually on Sundays.
1. Go to "League" > "Statistics" > "Sortable Stats".
2. Make sure there are no filters.
3.The position should be set to "all players", the scope should be "all levels", and the split should be "regular season".
4. Make sure it's a scroll bar and not paginated.
5. Then hit report, and "Write report to disk"
6. This opens a new tab on your browser
7. Once the new tab has FINISHED LOADING (important), you can right click and hit "Save as…"
8. Save the file (should save as html) as year_league_Ovr (example 2051 in Perfect League 400 is

2051_P400_Ovr.html). Make sure the league is correct - Diamond is D401, Gold is G500, and so on.
9. Apply the split "versus Left" and repeat steps 5-8, replacing the "Ovr" in the file name with "vL"
10. Apply the split "versus Right" and repeat the steps, using "vR".
Now we need to convert these to csv files. I made a basic file that reads everything and tries to grab the main table and convert it (in the [codebase](#), run 'python parse_new_data.py'). Or you can open the html file in Excel, but you'll need to delete any row that's not a header row or the data. Then save the file as a csv.

We take the vL and vR data so that we can calculate how batters perform against one side. This allows us to figure out how important splits are. Someone with 85 vL home run power will not hit the same number of home runs as someone with 85 vR home run power. This is because the average skill of the opposing pitcher will be different. By getting both sides of the data we can calculate a lot of stuff as we'll see later. From the data we can also parse out pitchers used as starters only and relievers only. We are unable to parse out the stats of a pitcher who started and relieved without lots of trouble. In most of the current meta, this is more rare during the regular season.

Tourneys are different. Once a tourney finishes you can gather the stats from it. Unfortunately, OOTP does not give us access to anything other than the "overall" split. There is no lefty and righty data. You can use the same process to gather the html files, with the naming scheme - tourney type_# of teams"T"last four digits. For example a cap bronze tourney with 32 teams and id 7105372 should be CB_32T5372.html. Following this will allow us to separate by tourney if we want to perform analysis.

In the [sample codebase](#), I've included some sample data pulled recently. We'll be performing analysis on it soon. But first, a quick stop to talk about BABIP, the hidden stat.

Calculating BABIP

TL;DR The BABIP formula is weird, but we can get within 3 points of the real value for >90% of the ratings.

The BABIP hitting rating is a hidden stat for OOTP. It is one of the few ratings not publicly available in the database. The others are some batter endurance stats, the GB rate for hitters, and hit by pitch ratings. If we can get this information, we have an edge over players who don't have it.

BABIP, POW, and AvK combine to form the CON rating. In OOTP 2021, the team changed the formula by making power uncapped. There were no public (at least 1 private) formulas to calculate BABIP from CON. I'm going to talk about how I set out to reverse engineering the BABIP rating. This will be reproducible

Most formulas in OOTP are linear - meaning we don't have to worry about values being squared or anything. Often, OOTP will introduce "breakpoints" in their formulas. For instance, the formula to calculate CON might be different when power is below 50 vs above 50. Through user reports and the edit tab when in commissioner mode on the main game this seems to be true. To figure out these formulas, we can gather hundreds of points of data and reverse engineer. Breakpoints from reports seem backed up by data:

low (< 13) and middle (50) for AvK
low (< 13), middle (50), and high (110) for POW
Low (13) and middle (50) for BABIP (the low for BABIP might not exist, I couldn't decide)

There doesn't seem to be a high breakpoint for AvK or BABIP currently. But, we've also rarely/never seen players with >110 in those stats, so it's something we don't need to worry about right now.

We want to gather data for this, we'll need to record 4 columns in a csv - CON, BABIP, POW, and AvK (I entered data in that order). For each input of a set of BABIP/POW/AvK we'll record the CON rating on the card. There are a couple caveats right now: 1. We will only look at the CON on the public non-editor part of the card (profile tab). We will not use the CON in the editor because there isn't always a 1:1 translation between the two. 2. We use the card numbers (profile tab) for BABIP/POW/AvK double-checking them with the editor number. The 2nd caveat is important, because the OOTP editor tab has a bug in it. When you edit the power columns, occasionally, it can misread the power rating when calculating the CON rating. This is easy to "fix" by editing another rating. I usually highlight the AvK rating and re-type it. To convert between editor number and public number I use the "odd" numbers in the editor. $(X + 1)/2 = Y$, where X is the editor number and Y is the public number. Steps to gather data:
Start a game with commissioner mode on and go to the "Editor" tab.
Edit the vL BABIP/POW/AvK stats to be different odd numbers (33, 55, 77 would work) and record the contact rating
Choose 1 factor to control at a time - start with babip
Send babip to the lowest (5, because things get wonky at mega-low values), which means a 3 on the public side.
Record the set of ratings
Add 4 to that 5 babip in the editor for a .
Record a new set of ratings where the public side of BABIP is now a 5.
Keep adding 4 and recording each set of data.
When you do power, the CON moves a bit slower so I usually jump by 8 instead of 4.
For BABIP and AvK I go as high as 221 in the editor. For power I go as high as 193 usually.

After doing this, you've collected some BABIP data for the low/high babip with middle AvK (between 12-50) and the middle-low power (between 12-50). You need to record variations of BABIP for: 1. low/low AvK/POW 2. low/middle-low 3. low middle-high 4. low/high 5. middle/low 6. middle/middle-high 7. middle/high 8. high/low 9. high/middle-low 10. high/middle-high, and 11. high/high. And then repeat the process for variations of POW and AvK. When I did this, I pulled 685 distinct sets of ratings, which you can see here. This will take a while, but is useful to know if you think the BABIP formula has changed.

Now we can break this data down and run regressions for each "category". We should be seeing r^2 values of >0.99 here. In the end I create a large matrix for low BABIP and high BABIP. Because we're calculating from CON, we don't know whether the BABIP is low or high. We'll have to calculate both and choose the one that seems correct. Generally one of the values is "invalid" - either the low BABIP value is negative or > 50, or the high BABIP value is < 51 or > 115 or so. If both are "reasonable" then I usually default to choosing the

high value so I don't under-calculate.

Ok, we have some formulas. Let's incorporate BABIP into the spreadsheet we're generating. We can generate a matrix to use, then add it into the cards. Check the new branch.

Thanks to YourKidnies for alerting me the old formula no longer worked.

Last edited by scipper; 03-16-2021 at 12:16 AM. Reason: bolding tl;dr

QUOTE

03-16-2021, 12:05 AM   #3

scipper
Bat Boy

Join Date: Nov 2018
Posts: 19

Baseball Stats Analysis

Basic Analysis
TL;DR You need many leagues of data for valid analysis, each from the same league level. Lots of code here to parse leagues.

Now we can talk about the main way of calculating new stats from what we have. This is where we get into heavier statistics. I'm not a statistician. I am an enthusiast at most. But this is my best attempt to give you a base. People who understand statistics feel free to ignore most of my advice.

First, we'll talk about data. When performing analysis, the more data we have, the better. A single season is worthless to try to perform analysis on because of how OOTP calculates stats. Each league is somewhat distinct and we need many to analyze. I'd recommend 4-5 as a minimum of leagues before metrics start to make sense, although I prefer 10+. Data from different levels of leagues is not preferred. This is because OOTP bases things around the average level of the league. Bronze leagues will have a lower average level than Diamond leagues. Also, league characteristics change over time. 10 Diamond leagues from 10 weeks ago will look different than 10 Diamond leagues from this week. When possible, separate any analysis by leagues. I would recommend a Diamond level analysis and a separate Perfect level analysis (assuming only 1 perfect level league).

When analyzing tournaments, you need an even larger amount of tourney pulls. This is because tourneys have a much lower number of events (because teams get knocked out). Unfortunately, since tourneys do not allow vL and vR pulls (only the Ovr pull), our analysis will be weaker, too. The benefit of tourneys is that players don't gain defensive experience. We can be more exact about who is playing what position. Remember to wait until a tournament is complete to pull data. I recommend at least 10-15 as a minimum before performing analysis.

Most analysis can be performed with simple linear regressions (sometimes non-negative linear regressions).

That's how I accomplish all my research. If you feel drawn to random forests, neural networks, or ridge regressions feel free to work with those. All the code I'll provide here will be using linear regressions (with some outliers removed). I remove outliers by using a stat called Cook's Distance. I found this page very helpful in using it. My method generally goes: 1. calculate a regression 2. figure out which individual pieces of data move our model too much 3. remove those with Cook's Distance 4. re-run a regression. We want to isolate variables as much as possible. Rather than trying to see what (con/gap/pow/eye/avk -> .wOBA) looks like, we'd rather get (eye -> walks) and calculate wOBA later. I'll cover this in later posts, but, generally, we can break down all stats.

Let's start analysis with something very basic. OOTP bases its seasons on 2010, but the 2010 league constants aren't great for calculating wOBA. With the league data, we can calculate some better wOBA constants. This should hopefully help us better rank players according to what OOTP will show us. Later on we can talk about developing our own wOBA based on data, not using OOTP constants. Anyway, to understand wOBA, and most any WAR stat, I won't re-explain here - please just read the truly amazing articles from Fangraphs instead (and consider subscribing!).

We need to read the league statistics we pulled earlier. Then we run a regression, pulling out the correct linear weights. We're going to calculate some r squared's here just as a check to be sure our wOBA factor formulas make sense. We're expecting values of 0.999 or so.

If you look at the commits for this part, it's a lot of code, but it lays the foundation for a bunch of later changes.

League Stats with no analysis
TL;DR Looking at many leagues of combined statistics is useful for identifying good players.

Let's look at what we can do without running any regressions first. We can generate the "average" stats of a player based ONLY on performance in leagues. Here, the important part is choosing which leagues to include. Again, don't mix bronze leagues and perfect leagues here. What we're looking for is consistent performance over many PA or BF.

We want to add up all the stats for vL/vR for the players, then calculate statistics based on the vL/vR breakdown. We can use our basic stats reading we calculated in the part before to do all the heavy lifting. For simplicity, we'll also pull in some basic ratings to this spreadsheet so we can filter on those, too. Check out the updated new sheet here.

Last edited by scipper; 03-16-2021 at 12:16 AM. Reason: bolding tl;dr

03-16-2021, 12:07 AM   #4

[scipper](https://forums.ootpdevelopments.com)

Bat Boy

Join Date: Nov 2018
Posts: 19

Hitting

Order of Batting Operations

TL;DR Hitting happens in a set order: 1. HBP 2. BB 3. SO 4. HR 5. Hit 6. XBH, and low ratings earlier in the order mean you have less chances later on.

Now we can move on to actually running regressions on batting data. The important part to realize here is the mechanics of how OOTP determines a batting outcome. OOTP does not calculate on a per-pitch basis but rather a plate appearance basis. OOTP calculates different batting outcomes (walk, hit, home run) in a certain order! Let's talk about how we can handle this.

At a basic level we could give our regression machine the factors (contact, gap, power, eye, avoid k) and output wOBA. This gives a vague sense of who is good, but doesn't do great. A great thing to do here is to calculate based on whether the pitcher was a lefty or a righty. This will get us using the (contact vL, gap vL, …) stats to calculate a wOBA vL and wOBA vR. Next is taking into account handed-ness. A righty batter versus a lefty pitcher with the same ratings as a lefty batter versus a lefty pitcher. I also break out switch hitters here, because OOTP seems to treat them differently. That's 6 different regressions so far we're performing: 1. lefty batter vs lefty pitcher 2. righty batter vs lefty pitcher 3. switch batter vs lefty pitcher 4. lefty batter vs righty pitcher 5. right batter vs righty pitcher 6. switch batter vs righty pitcher. There aren't many switch hitters in the game. Make sure you have enough different switch hitters to start drawing valid conclusions. At this point we could feel ok about our data. But we want more.

Based on research, the breakdown of an OOTP at bat goes like this:
Did the plate appearance result in a hit by pitch (batter HBP rate vs. pitcher HBP rate)
Did the plate appearance result in a walk (batter EYE vs. pitcher CON)
Did the plate appearance result in a strikeout (batter AvK vs. pitcher STU)
Did the plate appearance result in a home run (batter HR vs. pitcher MOV) [edit: see below comments, there is some debate whether this is calculated at the same time as strikeout. My code assumes not, but I may be wrong]
Did the plate appearance result in a hit that wasn't a home run (batter BABIP vs. defense)
If there was a hit, did it result in extra bases (batter GAP)
If there were extra bases, was it a double or a triple (batter SPEED)

A batter who walks more will have less chances to strike out. Take 2 lefty batters vs a lefty pitcher. Lefty batter A has an Eye rating of 70 and an Avoid K's rating of 50. Lefty batter B has an Eye rating of 30 and an Avoid K's rating of 50. They will not get the same number of strikeouts. Each batter's eye rating is compared to the pitcher's Control rating. In this example over 600 plate appearances, batter A might rack up 80 walks, while batter B might rack up 35 walks. This means that batter A has 520 (600 - 80) chances for a strikeout, while batter B has 565 (600 - 35) chances. If they both strike out at the same rate (10%), then batter A has 52 strikeouts and batter B has ~56. This works for all the steps. A low-Eye, low-AvK, high-Pow, high-BABIP player might never get the chance to use their skills because they strike out so much. A rule of thumb is that both Eye and AvK ratings have minimums. In high level leagues, player's who fall below them will not

perform well enough in their other ratings to make do. The minimums get higher as you go up in league levels.

Aside: These steps are verifiable by taking out each piece and looking at graphs of eye / pa vs walks and so on. This lets us verify that these steps are in the correct order.

We can calculate individual outcomes of wOBA instead of wOBA itself. We calculate walks vs lefties, singles, doubles, and so on. With the wOBA factors we found earlier, we find a projected wOBA. As a note: the DH position seems to take a small hitting penalty. They will perform worse than if they were in a non-DH position.

In the future, we can incorporate this into a WAR stat or combine the wOBA vL with the wOBA vR in different ways. Calculating this is enough for today. Here is the updated code at this point.

Full time starter vs. vL starter vs. vR starter
TL;DR You can protect a player if they have very lopsided splits. Keep in mind they'll still face wrong-sided pitchers.

Let's determine what kind of breakdown of pitchers a batter is likely to face. This breaks down into two distinct categories - catchers and everyone else. This is because catchers get tired faster so they're more likely to face a weird split of pitchers vL/vR than a fielder. As part of this, we'll also calculate how many games a player is likely to start. This gives us a sense of what we can expect for our counting stats when starting someone vL vs full-time.

To calculate these kinds of splits, we calculate games started and the % of PA's they face a right handed pitcher. To analyse this we don't look at the league level - we look at individual teams. We have to use some heuristics to measure what players were *probably* started vL/vR or as full-time. You can look at how I attempted this in the 'calculate_splits.py' file. Doing this we add in new statistics to our analysis and stats sheets, based on our vL/vR data. See updated code here. I've also thrown in some basic pitcher splits although I don't use them yet.

Note: all players have a hidden rating not on the card for the rate at which they fatigue. you see your catcher starting 5 less games than usual, they may fatigue faster than normal. This is per player (so all Legend Ted Williams will have the same fatigue rate).

HBP rate
TL;DR HBP is a hidden rate but can make a noticeable difference in your players. Have a look at a player's HBP rates because they should be very consistent even when other stats aren't.

The rate a batter gets hit by a pitch (and the rate at which a pitcher hits a player, too) is set by an inner hidden rate. This rate doesn't change and varies between high level and low level cards. Even great cards can have a bad HBP rate and bad cards can have a good HBP rate. This adds on to a player's walks without

having to deal with a pitcher's control rating. Even better, it is the first thing calculated in most at bats (unless it's an IBB). Since this rating is so consistent, we don't have to project it. As long as we've seen a player take a certain amount of plate appearances, their HBP rate won't change. Unfortunately, since it's hidden, we can only know it for player's we've already seen. Any new players we just have to wait for data on. See how we can calculate it here. I've found that a batter's HBP rate usually is between 3-12. Major credit to OOTP user Sipimi for bringing this to my attention. We'll explore the pitcher side of this stat later.

Last edited by scipper; 03-16-2021 at 10:47 PM. Reason: bolding tl;dr, editing for comment.

QUOTE

03-16-2021, 12:08 AM   #5

scipper
Bat Boy

Join Date: Nov 2018
Posts: 19

Fielding

TL;DR Range is king. We'll dive into more catcher effects in the next article (which will make catcher the most valuable). For now, SS and CF data matter a ton.

Defense is one of the hardest stats to figure out and find data for. The ideal thing we'd like to calculate is how many runs a player saved. There's a problem - each hitter can be compared to other hitters. We should only compare each defender to other defenders at the same position. Using regular league pulls, this is hard to figure out. We can cheat it by using tournament data.

In tournaments, if a player has only 1 position with experience to start, and they have innings played in the field, they played at that position. Rarely, they'll play at a position where they have no experience, but we'll have to ignore that. If we download tourney data. We can then analyze it like we do batters, We will only look at single-position players. Finding these, we can try to calculate defensive stats for each player. Important caveat: this requires a lot of tournament data, and of the same level. Mixing levels could throw off our averages. I recommend pulling 30-40 different individual open tourneys with no modifications (historical/live/…). Pull from around the same time (so the meta doesn't change much). When you enter one of these tourneys, I would try to make sure each player you start is a 1 position player. Especially at positions like LF/RF/3B where people start multi-position players. Figuring out how to better categorize positions or a better way of pulling stats would give a major edge.

We use individual ratings because they're worth more than the main defensive rating. Generally 1 point of range is worth more than 1 point of error rating. We want to figure out how that maps to run outcomes. While a player is training at a position any formula we generate won't map to the expected outcome. There have been anecdotes of players slumping while training a new position. But, I don't have any data to back it up.

Given this data, we can look at 2 things. The first is ZR per inning. ZR is good because it is already translated to runs saved and adjusted per position. If we know the number of runs per win, we can translate it to how many wins a player added to. The downside is that the way OOTP calculates ZR is unknown. They

may not actually be using the best way to calculate runs saved. The other way is to look at the actual ball in zone data. OOTP tells us: 1. How many total balls passed through a defenders zone. 2. How hard it was to get to the ball (routine play, likely, even, unlikely, very unlikely, and impossible) 3. Whether they made it to the ball. Using these combined stats, we can calculate a play%. We can also figure out what % of times a ball entered in the zone will generate about how many outs. This is all straightforward so far. Unfortunately, there have been reports that range affects total balls in the zone. This means we can't expect the same amount of balls in the zone per player - it depends on their range. By doing this play% gets thrown off because a higher range player could have more unlikely balls in zone. They could be converting "impossible" zone balls to "unlikely". To calculate everything I went and calculated a normalized play%. After, I also calculated total balls in zone. Finally, I determined how many outs above average based on this. With these outs above average (per 162 games) we can translate this into WAR if we know outs per run.

No positional adjustment for WAR yet, so the raw data looks like RF/LF is worth more than CF. This is because the standards are higher for CF, so the replacement cost is higher. Generally we'd expect CF/SS to be the most important WAR-wise worth about 2-3 wins over competition. At high levels everyone is starting some of the best defensive players.

For now, we'll ignore the effect of catchers on pitchers.

See updated code here, especially the calculate_defensive_stats.py file. There's a lot of intricacies of logic here that you may decide to remove or not.

Catcher Defense
TL;DR Catchers have a massive effect on CERA. They're the most important defensive position in the game.

In the previous article, we looked at how players affect balls in play. 7 out of the 9 positions have their only defensive effect here. Catchers do more than that, though. Catchers affect how well a pitcher pitches. We need to look at how well catchers turn hits into outs and walks into strikes. We do this by focusing on CERA. If we do a regression on CERA, in OOTP 21 (this changes per version), about 1 point of C ability is worth 0.01 CERA. Over a full season this translates to 100 Catcher ability is worth about 10 WAR! That's an insane number. We can't be perfect because the best teams play the best pitchers and best catchers together. They'll play teams who have neither. In general we can see how valuable this is.

The other main stat that catchers can affect is stolen bases. Not only by throwing runners out, but deterring runners from even attempting a stolen base in the first place. A higher catcher arm ability leads to a small increase in runners thrown out, but a large decrease in attempts. We can calculate how much a catcher "deters" stolen bases too.

Catchers value is defensive so remember to grab a strong defensive catcher first. See updated code calculating this here.

QUOTE

03-16-2021, 12:10 AM

scipper
Bat Boy

Join Date: Nov 2018
Posts: 19

Running

TL;DR Running stats are uncapped, so you can take advantage of this. High stealing/high speed players can mess with other teams.

Let's talk about how many hits there are at the end of the season. During the first couple games (of PT) of a season, OOTP figures out the average stats of those playing. From those stats, it calculates the correct rate to match the hits total of the 2010 MLB season. It then sets those rates and keeps them for the rest of the season. This means that the league will produce within about 5% the 2010 MLB season totals. The same number of hits, home runs, walks, … (Discussion about it here. Generally QuantaCondor knows a lot about this. Others who have talked about this on the forums and who are much more knowledgeable than me - Syd Thrift and RonCo). Hits, home runs, walks, all match season totals, but a couple stats don't… steals being one of them. This means that steals don't have a limit. With the other stats, you're battling others for a slice of the pie, but with steals, there is an unlimited pie. In some perfect leagues there were 7000-8000 steals in a season. The highest amount in the last 20 years was in the 3000's. This is great for people who know how to get a steal. Let's talk about the ratings - speed is how often they'll attempt a steal, but does not affect success. The steal rating is how often they'll succeed at a steal, but not how often they attempt. So a high-speed/low-stealing player will attempt a ton of steals and get caught a lot. A low-speed/high-stealing player will not attempt many but succeed often. You should edit your sliders so high-success stealers to push their luck more (and vice versa). Baserunning affects how they'll do along the base paths outside of steal attempts.

This is the code I've worked on the least - especially UBR. I don't even try to calculate wGDP. Improving this part is a good area for improvement. See new stuff here.

QUOTE

03-16-2021, 12:11 AM

scipper
Bat Boy

Join Date: Nov 2018
Posts: 19

Putting it all together - a WAR stat

TL;DR We try to put the hitting, fielding, and running stats together to create a WAR stat. See sWeAR in the sheet.

With all the offensive stats we've calculated so far, we can try to calculate a WAR stat now. Following Fangraph's guide we can put everything together. But hey, we need some constants to help us know how to convert wOBA to runs and all that. We can re-use our wOBA constants and calculate what the wSB stats, outs per run stats, and runs per win stats are for OOTP. Or, we could calculate our own constants using this

article. I'm not currently using it in a major stat but I added the code as an example in the linear_weights folder.

I adjust the defensive ratings to account for everything. You may need to change the adjustments yourself. Full code for sWeAR (scipper's wins expected above replacement) is here.

QUOTE

scipper
Bat Boy

Join Date: Nov 2018
Posts: 19

Pitching

TL;DR Remember the order of operations here - 1. HBP 2. Walk 3. K's 4. Home runs

Now for the pitching side. We handle this like the hitting part. OOTP assumes pitching is DIPS. Pitchers don't affect BABIP much except groundball rate. This means we can look at pitching pretty much only as 4 outcomes: 1. Walk 2. Strikeout 3. Home run 4. Ball in play (where the defense takes over). Pitchers affect the first 3 outcomes. The defense affects the 4th.

First, OOTP determines whether the batter is hit by pitch. This is based on a hidden rating. After that OOTP determines if there was a walk.This checks the pitcher's control rating vs. the batter's eye rating. After, the game checks for a strikeout, checking the pitcher's stuff rating vs. the batter's avoid K rating. Then, there's contact, so the game checks the movement rating vs. the power rating to see if there was a home run. Finally, there must be a ball in play. OOTP checks both the gb% for the pitcher and the batting profile for the pitcher. These are both at least semi-hidden ratings. If the defense doesn't make an out, then it's a hit.

In OOTP the control rating is the premium one. A minimum movement generally expected, and a higher minimum stuff rating. Handedness matters here.

Based on the same way we determine hitting, we can run similar code to calculate FIP for pitchers. Here's the output.

HBP rate
TL;DR Double-check pitcher's HBP rates because they can be up to 25 extra batters walked a year.

I talked about HBP rate for batter's but the more important one is the HBP rate for pitchers. Batters vary a little (from 3-12 walks a year or so), but pitchers can vary a lot (from 3-25ish). This is worth several points of control and can make a huge difference. When your pitcher is playing, double check that they aren't underperforming because of HBP. HBP stats already in the previous post's code.

Edit: I forgot to mention wild pitches. Pitchers have a hidden wild pitch rate, and that affects how many bases

are stolen on them too. Haven't included it in code, but a useful extension if you're looking at that data specifically.

Last edited by scipper; 03-16-2021 at 04:27 AM.

QUOTE

03-16-2021, 12:14 AM   #9

scipper
Bat Boy

Join Date: Nov 2018
Posts: 19

Cleaning up
Where to go from here
I've put a lot of code and talk in here, but let's talk where we can improve.

First, let's talk about the overall. You could start weighting data by "pa" for batters or "bf" for pitchers. You could also start mixing the projections and the stats you pull. If you use the projections as a seed and adjust them based on actual stats shown. If you give your initial projections a power of 3000pa, then start incorporating actual data, you can move to better data.

For batters, the places with the lowest hanging fruit are defense and running. Running has the least work done on it, although it seems to be worth the least (relative to hitting and defense). Defense has some very wonky stats and needs special investigating. Much more data needs to be pulled. The stats should be useful, but you should train your instincts by comparing them with real data. This will help you identify the differences between projection and real stats. If the batted ball profile data becomes available, make use of it in the hitting statistics. It should be very helpful to be even more accurate.

For pitchers, there is a big gulf for things. If we can predict walks, homeruns, and strikeouts, then we know how many balls in play there are. Given that, we could come up with an estimation for ERA (assuming average defense). This is more useful. I've had a lot of trouble estimating the number of innings pitched, too. Estimating that would give an edge. One thing I did during the OOTP 21 season that I haven't included is record SP-as-RP stuff ratings. It would be great if OOTP would publish every pitcher's stuff rating as an SP and stuff rating as an RP. This would allow us to see which SP transition to RP and rate them for RP FIP. This past year I hand-curated a bunch of ratings, but have not included that in the code. I don't think it would be that hard to put in place, you would change the "CID" field to be different from "t_CID". Add an -rp if it's SP-as-RP or -sp if it's RP-as-SP. I've left some comments in the code where you'd need to edit to get things to work. I also have not dealt with the fielding part of pitching. Currently I do no analysis around the hold rating for pitchers or how often they have to field balls. A more complete analysis would also punish pitchers for a bad hold rating.

For tourneys, you can change the current code to generate tourney-specific stats files. This allows you to see which players are performing the best in specific types of tourneys.

Remember to look at park factors. Adjust your team to fit. It wouldn't be that hard to include park factors in your hitting ratings. In pitching ratings that may be harder. Lots of people use openers to try and catch your lineup with the wrong handedness. Build your team to guard against that strategy if you can.

This covers most of the research I've done over the past year. Good luck!

QUOTE

03-16-2021, 03:00 AM

#10

QuantaCondor
All Star Reserve

Join Date: Nov 2018
Posts: 547

Amazing, ridiculously detailed post. This should be standard reading (if not stickied) for anyone looking to dive deeper into how PT works.

I'll add a few general comments that I think are worth saying, just as a supplement from another PT modelmaker's point of view. Overall, many assumptions stated here and more broadly in older posts in the community are about how the PT21 engine works (and in some cases, about earlier games). There's no guarantee that one single thing stays the same from year to year. That's why the most important takeaway here is the overall strategy for gathering data and analyzing it, rather than any specific assumption. This is especially true for statements like "this particular stat is important" or "the breakpoints are here".

For example, avoidK wasn't nearly as important in PT20. Control wasn't nearly as strongly scaled in PT20. OF defense was much worse in PT21 vs PT20, and 2B defense was much better. Catcher defense changed radically. And when you switch out the eras or metas for something else, something which I anticipate will be relevant in PT22, you need to basically re-evaluate many of your scale factors and assumptions to see what holds and what changes. Even just within the same game, moving from base game to standard tournaments changes how you have to view players by a lot. The best players know how the different environments affect their assumptions, which is an intuition you develop by analyzing these different environments and seeing what changes and what stays the same.

Another general point I'll make is that I have found weighing your regressions by PA and IP make a ton of difference in terms of how much data you need to assess a particular environment. Some stats need more data, like HR%, but things that happen more frequently like BB%, K%, and even BABIP scaling converge much more quickly. This is useful especially in things like perfect or certain tournament formats where you don't have the luxury of many weeks of data and/or the format meta shifts quickly and you want to be proactive about solving it.

The last thing I'll mention is that the most important feature of a good OOTP model is validation. You can slam whatever coefficients into your spreadsheet or script you want, but comparing the projections to what you actually see (including at extremes!) is frequently where you learn the most about the game. It's also enlightening to compare your analysis to that of other players, to test your assumptions, and constantly question if what you're doing is really the best way to do it.

Again, awesome post by scipper here. Many of the comments are true regardless of the PT version you're on, so hopefully it can be one of those posts people link to even beyond PT22.
_____

Former leader of BFF, the definitive competitive PT group for F2P players. DM for info

F2P + restrictions. First F2P winner of PT21 Perfect League

F2P + restrictions. New team -> PT title in 8 weeks

QUOTE

03-16-2021, 03:56 AM   #11

BennytheKid
Bat Boy

Join Date: Jun 2019
Posts: 3

Some of you might know this name but this is Your Kidnies.

This man is the entire reason my Sheet exists. If you ever wanna learn how to do what I did this is how you do it right here. Just amazing stuff. Love you buddy

QUOTE

03-16-2021, 12:34 PM   #12

dbqs
Minors (Triple A)

Wish the "Thanks" button was still here so I could smash it

Join Date: Aug 2017
Location: Comiskey
Posts: 237

QUOTE

03-16-2021, 01:14 PM   #13

dvanhout
Bat Boy

Join Date: Jun 2019
Posts: 8
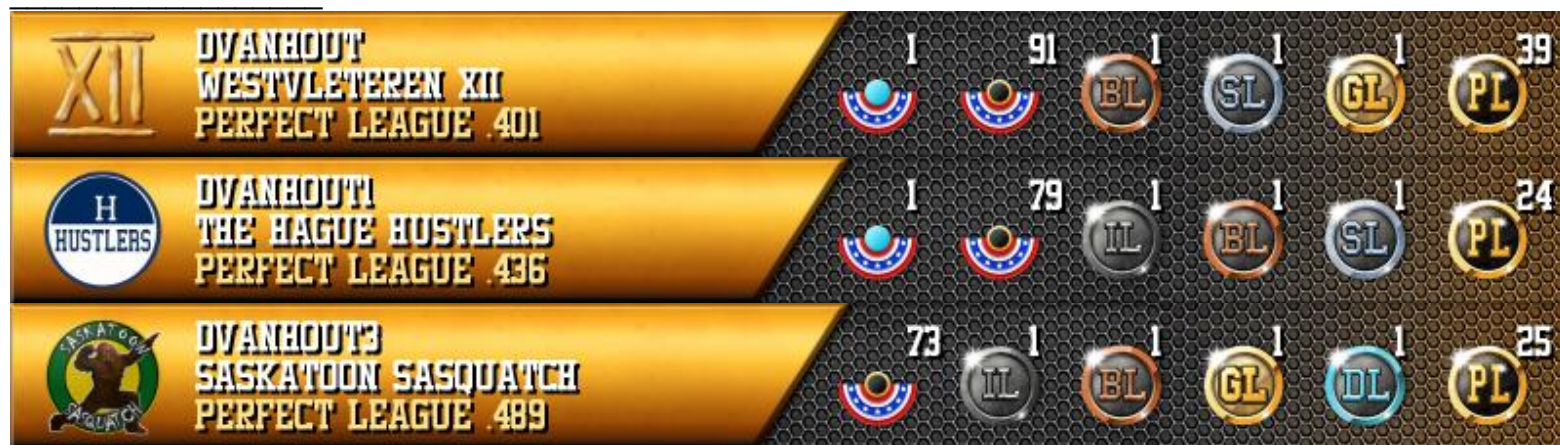
This is good stuff. Appreciate the detailed info.

QUOTE

03-16-2021, 01:27 PM   #14

Crash
Minors (Rookie Ball)

Join Date: Nov 2018
Posts: 45

Thank you for your information. This is awesome.

QUOTE

allenciox
All Star Reserve

Join Date: Dec 2015
Posts: 560

Excellent, some questions and notes.

Thanks for the post, as QuantaCondor says, it should be required reading for those wanting to understand how OOTP engine works.

So, a few questions:

1. How do you know that strike outs are determined before home runs? In the analyses I have done I haven't detected any conditional dependence for home runs on strike outs, or vice versa (as opposed to other things): for example, there being a higher r-squared value for predicting home run percentage with strike outs removed from AB as opposed to AB directly. Yet this also kind of makes sense as it would explain part of the reason why AvoidK has such a big relationship to wOBA and POW has such a small effect (less than GAP, even) at the highest levels of Perfect Team. What is the evidence supporting this?

2. Are you sure that the following quote in your post is accurate? The reason I ask is that I thought the LTMs (league total modifiers) were calculated by simming a portion of the schedule before the season (or tournament) starts, and then adjusting the results based on those sims. I thought this was part of the reason there is a delay after all teams have signed up before a tournament starts, so that this simulation can occur, and that this is also the source of the "preseason predictions" for standard leagues. But I could be wrong.

> Quote:
>
> Originally Posted by **scipper** ⊳
> *TL;DR OOTP figures out the average stats of those playing. From those stats, it calculates the correct rate to match the hits total of the 2010 MLB season. It then sets those rates and keeps them for the rest of the season.*

One point I will add to your excellent description of the advantages and disadvantages of tournaments relative to perfect team seasons for data analysis:

In tournaments other than OPEN, you have a restricted set of cards that are typically played; this means that predicting what is successful is going to have more "staying" power than it is for perfect team seasons, unless additional powerful cards are released specific to that tournament structure.

ALLENCIOX CARY SOUTHPAWS-OL DIAMOND LEAGUE .350	1	1	16	7		2 GL		1 DL
MILOWINDBY WESTON POINTE CASPERS-OL GOLD LEAGUE .356	1	1	1	23	5		1 SL	
HOMEOFTHEBRAVES CARNEGIE HALL PERFECT PITCH-… DIAMOND LEAGUE .372	1	1	6	28		2 GL		2 DL

Lefties only need apply:

ALLENCIOX SOUTHERN SOUTHPAWS-OL PERFECT LEAGUE .435	1	2	2	88	1 DL	25 PL

Smallball team:

HOMEOFTHEBRAVES BATTERY POINT ENERGIZER BUNN… PERFECT LEAGUE .447	1	1	4	76	1 DL	9 PL

Power team turned righty team:

MILOWINDBY WESTON POINTE CASPERS-FTP PERFECT LEAGUE .430	1	4	1	74	1 DL	7 PL

QUOTE

03-16-2021, 10:37 PM #16

scipper
Bat Boy

Quote:

Join Date: Nov 2018
Posts: 19

> Originally Posted by **allenciox** ▷
> *1. How do you know that strike outs are determined before home runs? In the analyses I have done I haven't detected any conditional dependence for home runs on strike outs, or vice versa (as opposed to other things): for example, there being a higher r-squared value for predicting home run percentage with strike outs removed from AB as opposed to AB directly. Yet this also kind of makes sense as it would explain part of the reason why AvoidK has such a big relationship to wOBA and POW has such a small effect (less than GAP, even) at the highest levels of Perfect Team. What is the evidence supporting this?*

Anecdotally you're correct in my data, but I had mostly just followed other posts about the subject. It isn't a huge change to my r-squared's so I never could decide whether to keep including it or not.

Quote:

> Originally Posted by **allenciox** ▷
> *2. Are you sure that the following quote in your post is accurate? The reason I ask is that I thought the LTMs (league total modifiers) were calculated by simming a portion of the schedule before the season (or tournament) starts, and then adjusting the results based on those sims. I thought this was part of the reason there is a delay after all teams have signed up before a tournament starts, so that this simulation can occur, and that this is also the source of the "preseason predictions" for standard leagues. But I could be wrong.*

No you're correct - I was trying to describe that at a high level and didn't get it quite accurate.

QUOTE

03-17-2021, 12:50 AM   #17

QuantaCondor
All Star Reserve

Join Date: Nov 2018
Posts: 547

Quote:

> Originally Posted by **scipper** ▷
> *Anecdotally you're correct in my data, but I had mostly just followed other posts about the subject. It isn't a huge change to my r-squared's so I never could decide whether to keep including it or not.*

I also had HRs happening first (or at least not being obviously affected by K%) for what it's worth. Think about it: if Ks came first, it should have a BIG, noticeable impact on many low-avoidK sluggers. 30% K rate differences should make it immediately obvious. If HR came first, it should basically not have a huge effect.

But this is a good question for readers to answer for themselves using your analysis methods. Measurable, testable, and relatively simple.

_____
Former leader of BFF, the definitive competitive PT group for F2P players. DM for info

F2P + restrictions. First F2P winner of PT21 Perfect League

F2P + restrictions. New team -> PT title in 8 weeks

QUOTE

cavacom
Minors (Rookie Ball)

Join Date: Jan 2019
Location: Canada
Posts: 49

This is great and all but what about BALKS?

QUOTE

POST REPLY

BOOKMARKS

« Previous Thread | Next Thread »

POSTING RULES

Forum Jump
Perfect Team 22   Go

POSTING RULES

You **may not** post new threads
You **may not** post replies
You **may not** post attachments
You **may not** edit your posts

BB code is **On**
Smilies are **On**
[IMG] code is **On**
HTML code is **Off**

Forum Rules

All times are GMT -4. The time now is 01:39 PM.

-- OOTP 2017   Contact Us - Out of the Park Developments - Privacy Statement - Top

Major League and Minor League Baseball trademarks and copyrights are used with permission of Major
League Baseball. Visit MLB.com and MiLB.com.

Officially Licensed Product – MLB Players, Inc.

Out of the Park Baseball is a registered trademark of Out of the Park Developments GmbH & Co. KG

Google Play is a trademark of Google Inc.

Apple, iPhone, iPod touch and iPad are trademarks of Apple Inc., registered in the U.S. and other
countries.

COPYRIGHT © 2020 OUT OF THE PARK DEVELOPMENTS. ALL RIGHTS RESERVED.

Powered by vBulletin® Version 3.8.10
Copyright ©2000 - 2021, vBulletin Solutions, Inc.
Copyright © 2020 Out of the Park Developments