Ranking the MLB's Most "Deceptive" Pitchers

Jeanine Minnick
Fall 2020

"Deception" is the ability for a pitcher to disguise each of his pitches far into a pitch's trajectory, so the batter has to guess more than actually recognizing the pitch. Some pitchers are able to spin the ball more and create more break, some are able to throw much harder than others, and some may barely change their release point to create a more deceptive pitch arsenal. This is an attempt to use all of these pitch characteristics, along with specific at-bat results, to determine the most deceptive pitchers from 2019 and 2020.

This project will implement all Statcast data from 2019 to the present (as of 10/12/20). The publicly available data provided on Baseball Savant includes valuable metrics for calculating pitch deception, but there were a few notable pitch characteristics that are not included - horizontal break, vertical break, and spin direction.

The first step in solving this question is to use the collected Statcast data and create individual values to be added to the overall "deception" calculation. Here are the values calculated for each pitcher in the dataset (only pitchers with at least 500 pitches the dataset are included - any reference to "all pitchers" below refers to this subset of the 2019/2020 data):

- **Swinging strike percent** (swingingk_pct)- *deception can be indicated by a hitter's inability to make contact with a pitch*

- **Looking strike percent** (lookingk_pct) - *deception can be indicated by a hitter's inability to determine the pitch type of the pitch on it's way to the plate, as well as the movement on a pitch (did not include pitches that occurred on 3-0 counts, since it is a common pitch for batters to take - out of all 3-0 counts in this dataset, 56.2% were called strikes, 11.1% were swung at, and the rest were balls)*

- **Distance from average pitch release point** (rel_var_dbd]) - *a batter will have a more difficult time determining pitch type if a pitcher has a more consistent release location - this calculation was done on a day-by-day basis because pitchers are likely to make changes mechanically or change their starting point on the rubber on a game-by-game basis, therefore this value would not be as useful for all pitches over the two year time frame)*

- **Hit percent of pitches contacted** (hit_pct) - *the ability for a hitter to make solid contact on a ball is a good indicator of ability to deceive the hitter*

- **Variance of pitch speed** (speed_var) - *being less consistent with pitch speed is a common way to deceive hitters (idea is that a pitcher that throws fastballs 80% of the time will be more predictable, therefore less deceptive)*

- **Average Bauer Units** (bauer_avg) - *the higher the Bauer Unit, the more the pitch deceives the hitter. Calculated by dividing spin rate by pitch speed (metric from Driveline)*
    - *there were a number of pitchers in this dataset that did not employ a fastball of any kind - therefore their fastest pitch was used to calculate Bauer Units*

- **Total "area" of pitch break** (break_area) - *as highlighted in Rapsodo pitch metrics and a recent Rapsodo podcast, having more spread between the direction of pitch break can result in a more deceptive pitch arsenal (used average pitch acceleration in the x and z directions for each pitch type as an alternative since break was not in the dataset)*
    - *a few pitchers only have two pitches - line distance was used for break area (this does result in a generally lower break area for pitchers with only two pitches)*

After doing these calculations, it was determined that the deception metric would be informative if it is on an increasing scale, therefore a higher "deception" value is better. Since there are multiple variables calculated that are "better" if they hold a lower value, like hit percent of contacted pitches in the strike zone, the inverse of these corresponding values are used.

The first attempt of a deception calculation is shown below:

$$deception = swingingk\_pct + lookingk\_pct + \frac{1}{rel\_var\_dbd} + \frac{1}{hit\_pct} + speed\_var + bauer\_avg + break\_area$$

Looking at the cumulative results for each pitcher, if the above formula is used, each variable has a very different weight, therefore some of the variables are practically meaningless in the overall calculation (first three pitchers shown):

player_name	deception	swingingk_pct	lookingk_pct	rel_var_dbd_inv	hit_pct_inv	speed_var	bauer_avg	break_area
CC Sabathia	316.2	0.112	0.158	194.4	5.1	19.5	24.8	72.1
Fernando Rodney	252.4	0.106	0.164	78.4	5.6	27.5	22.2	118.6
Oliver Perez	235.5	0.114	0.197	27.7	5.8	47.1	24.0	130.6

It is likely that some of these variables are more meaningful to pitch deception to others, but for the scope of this question, each variable will be weighted equally. The average of each variable is calculated, and the corresponding linear weights to make each variable equal in weight are determined. The league average "deception" value will be 100 in order to easily compare a player to average.

The resulting deception formula is below:

$$deception = 126.0255 * swingingk\_pct + 89.09515 * lookingk\_pct + \frac{1}{10.588821 * rel\_var\_dbd} + \frac{2.422145}{hit\_pct} + \frac{speed\_var}{1.912593} + \frac{bauer\_avg}{1.718192} + \frac{break\_area}{10.51374}$$

Using this formula to calculate the deception values for each pitcher, we get the following results:

	player_name	deception	most influential variable in overall deception calculation
1	Taylor Rogers	152.9	release point
2	Stephen Strasburg	140.9	release point
3	Trevor Bauer	136.7	break area
4	Aroldis Chapman	136.1	break area
5	Grant Dayton	134.2	release point
6	Rowan Wick	133.4	speed variance
7	Kyle Crick	132.5	speed variance
8	Lucas Sims	129.7	break area
9	Rich Hill	129.6	speed variance
10	Alec Mills	129.4	speed variance
11	Walker Buehler	128.3	break area
12	Charlie Morton	128.1	speed variance
13	Liam Hendriks	127.8	release point
14	Shane Bieber	127.5	release point
15	Felipe Vazquez	126.4	break area
16	Nick Anderson	126.0	release point
17	Yu Darvish	125.9	break area
18	Mike Clevinger	125.4	speed variance
19	Brad Hand	123.9	release point
20	Gerrit Cole	123.5	break area

We would expect to see some of the names on this list, like Aroldis Chapman, but there is also a presence of a few names that are more surprising. A unique part about this formula is that it allows the pitchers to create a high overall deception value in many different ways. For example, Taylor Rogers has an extremely small range of where he releases, and has a very high looking strike percent value (22%, second highest among all pitchers), therefore these variables have a large effect on his overall deception value. Trevor Bauer is ranked highly on this list for a completely different reason - he has the highest break area out of all pitchers.

Although it is not included in the dataset, spin direction is a great way to create a deceptive pitch arsenal - take Tyler Glasnow for example. He throws a curveball with a very low amount of horizontal break, therefore it has the characteristics of a 12-6 curveball. His fastball matches that spin angle very closely, since it has a near vertical spin at 6:00. Pitchers can use pairing spin directions, like Glasnow's fastball - curveball combination, to deceive batters.

As a whole, the results of this calculation show to be promising. It would be valuable to determine the weights of each variable relative to the others, since certain metrics have a higher effect on deception. Collaborating with scouts and coaches to know more about what makes a pitcher deceptive would be a great starting point for making this linear weight decision. Also as stated before, if horizontal break, vertical break, and spin direction were included, as well as more information about the pitch trajectory itself (location of the ball at 50ft from the plate, for example), this calculation could be much improved.

Code:
https://github.com/jeaninem8/mlb_deception/blob/main/pitcher_deception.R