Does Umpiring Impact Pitchers' Performance Relative to Their Expected Statistics?

In this study, I will analyze a data set of all pitches thrown that were either a ball or a called strike in the 2021 season, focusing on which pitchers got "lucky." Luck will be defined as the percentage of pitches located on the borderline of the strike zone called a strike.

A better understanding of how luck-based factors influence the game could result in better player evaluation and the ability to adjust expect statistics for regression to average luck.

Defining Borderline

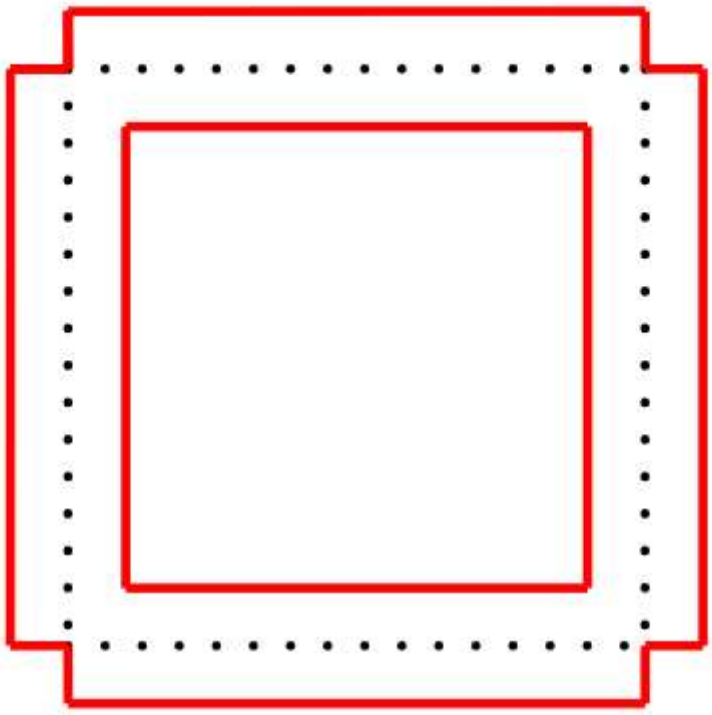
It helps to know what variables we are going to be working with. The Statcast variables for pitch location (Where the ball crossed the plate) are called plate_x and plate_z. plate_x is the horizontal position of the ball from the catcher's perspective. If plate_x = 0, the pitch was located directly on the center line of the plate. Unfortunately, these variables are not standard units of measurement. The accepted plate_x units for strike zone are -0.8 to 0.8.

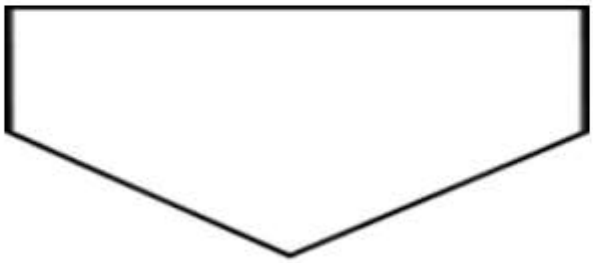
0.8 and -0.8 plate_x pitches look borderline as is. It stands to reason that pitches slightly closer to the middle of the plate are borderline, and pitches slightly further off the plate are also still borderline. "Slightly" isn't something that I can tell R, however. Let's first start by speaking in variables.

x_b represents "x boundary", which is the lower bound of borderline for left and right pitches. $x_b + \Delta$ represents the upper bound, with delta representing the range.

- Pitches in the ranges $-x_b - \Delta \leq x \leq -x_b$ and $\text{sz\_bot} \leq y \leq \text{sz\_top}$
- Pitches in the ranges $x_b \leq x \leq x_b + \Delta$ and $\text{sz\_bot} \leq y \leq \text{sz\_top}$
- Pitches in the ranges $-0.8 \leq x \leq 0.8$ and $\text{sz\_bot} - \Delta \leq y \leq \text{sz\_bot}$
- Pitches in the ranges $-0.8 \leq x \leq 0.8$ and $\text{sz\_top} \leq y \leq \text{sz\_top} + \Delta$

This image generally represents the area I've defined as borderline. The dotted line is the true strike zone and the red lines outline the borderline zone.





You might notice that this doesn't capture the corners of the zone. That's a sacrifice I'm making here for the sake of simplicity

To determine how narrow the inner red lines should be, I looked at pitches closest to the center line of the plate that were still called a ball. I watched videos of the pitch to ascertain how bad the call was as well as glean some of the context behind the pitch. I only looked at pitches between 2 and 3 plate_z units in order to filter out pitches that were called balls because they were vertically borderline pitches.

Unfortunately, these calls looked like outliers rather than genuinely borderline. The pitch closest to the middle that is still called a ball is an egregious call on an Aroldis Chapman slider, so egregious that "borderline" isn't a fitting description. The next closest pitch is one thrown to Jose Altuve. My guess is the umpire thought it was a high pitch, not an inside pitch, because of Jose Altuve's height. The third pitch is one where the catcher sets up inside to a right handed hitter, and the pitcher misses his spot outside, so the catcher had to dart his glove to the right to catch ball. I suppose you could argue this was unlucky for the pitcher - a pitch in the zone should be a strike - but

again doesn't seem like a "borderline" pitch to me. The fourth pitch looks like the Diamondbacks catcher was crossed up, not expecting to receive a curveball, meaning the call was less a matter of luck and rather miscommunication. The fifth pitch is also poorly received like the third pitch.

Baseball Savant Links to all described pitches, in order:

- https://baseballsavant.mlb.com/sporty-videos?playId=0698a942-8e00-4a77-b837-65d97ed3c161 (https://baseballsavant.mlb.com/sporty-videos?playId=0698a942-8e00-4a77-b837-65d97ed3c161) (Outlier call)
- https://baseballsavant.mlb.com/sporty-videos?playId=289d3d10-6f98-4d1a-8800-f59364efb246 (https://baseballsavant.mlb.com/sporty-videos?playId=289d3d10-6f98-4d1a-8800-f59364efb246) (Altuve's height)
- https://baseballsavant.mlb.com/sporty-videos?playId=9337b339-8c71-462d-97e2-8f2ad2fad9e2 (https://baseballsavant.mlb.com/sporty-videos?playId=9337b339-8c71-462d-97e2-8f2ad2fad9e2) (Catcher setup)
- https://baseballsavant.mlb.com/sporty-videos?playId=9d3a47fa-ec63-45c4-a7c6-7fb81658eed2 (https://baseballsavant.mlb.com/sporty-videos?playId=9d3a47fa-ec63-45c4-a7c6-7fb81658eed2) (Crossed up)
- https://baseballsavant.mlb.com/sporty-videos?playId=cecfbb45-30be-4186-86fe-a224c3c4ac9f (https://baseballsavant.mlb.com/sporty-videos?playId=cecfbb45-30be-4186-86fe-a224c3c4ac9f) (Catcher setup)

I'm struggling to find a good way to define borderline if I discount potential bounds by looking pitch-by-pitch. Instead, I choose to include pitches from $\pm$ 0.6 to 0.9. This is, admittedly, fairly arbitrary. This range is based on looking at examples of pitches with a plate_x of 0.6 and 0.9 and judging whether they appeared close to the edge of the plate. This range captures about a ball width fully inside the width of the plate, and a bit more than a ball length off of the plate.

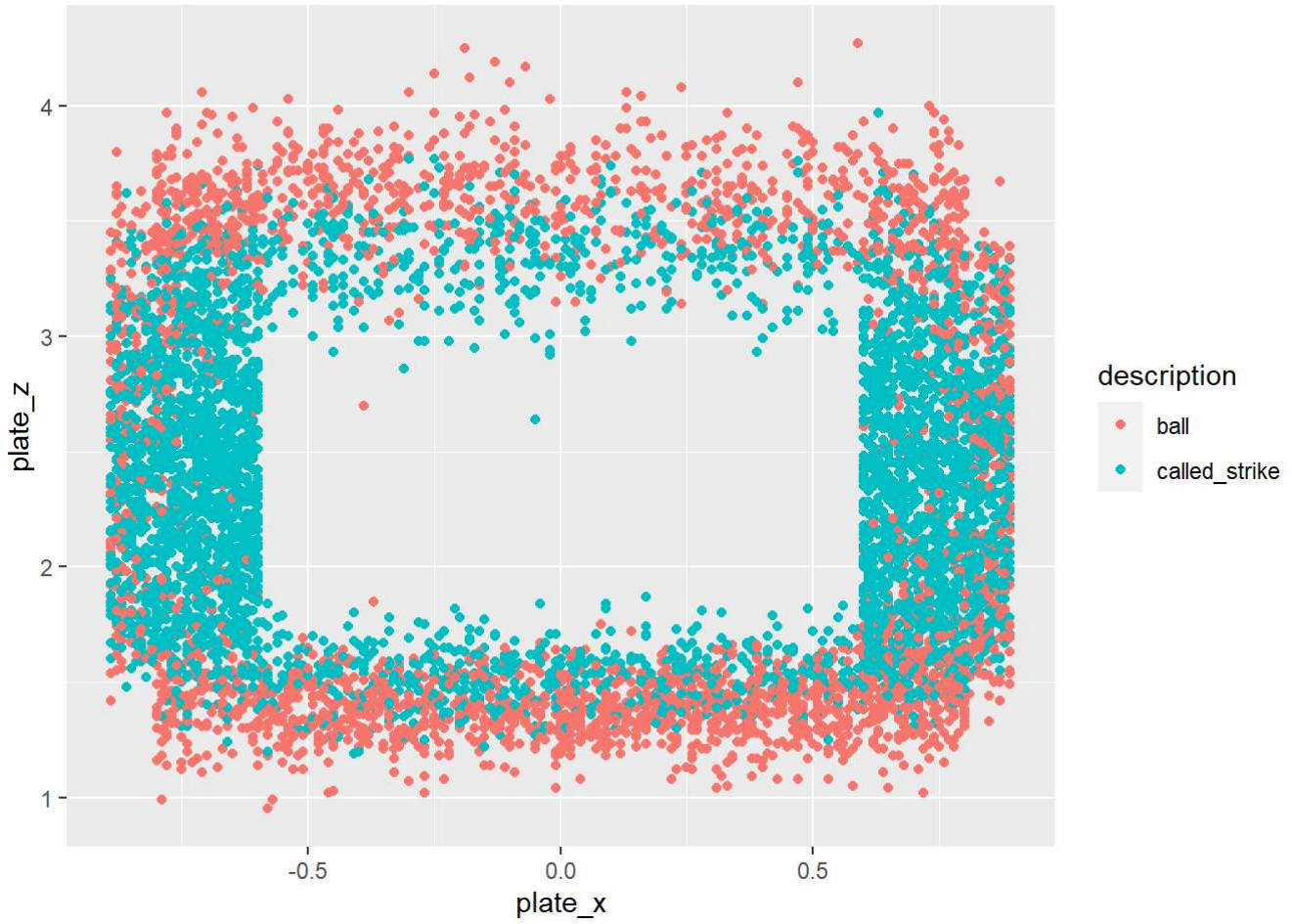
As a result, the upper and lower borderline will also have a width of 0.3 plate units. I don't know if sz_bot and sz_top refer to the lowest or highest possible pitch that might be called a strike or of it acts similarly to the edge of the plate. As a result, this definition might be even more arbitrary for vertically borderline pitches than it is for horizontally borderline pitches

This code filters all the pitches I have defined as borderline.

```
x_b <- 0.6
delta <- 0.3


Borderline_Pitches <- Pitches %>% filter({-x_b - delta <= plate_x & plate_x <= -x_b & sz_bot <=
 plate_z & plate_z <= sz_top} |
                   {x_b <= plate_x & plate_x <= x_b + delta & sz_bot <= plate_z & plate_z <= sz_
top} |
                   {-0.8 <= plate_x & plate_x <= 0.8 & sz_bot - delta <= plate_z & plate_z <= sz
_bot} |
                   {-0.8 <= plate_x & plate_x <= 0.8 & sz_top <= plate_z & plate_z <= sz_top + d
elta})
```

Let's get a sense of this dataset



With this definition of borderline, there are a total of 7,514 Pitches. 4,475 were strikes for a strike percentage of 0.5956

Borderline Analysis

player_name	ball	called_strike	total	strike_pct
Adam, Jason	0	4	4	1
Almonte, Yency	0	4	4	1
Arihara, Kohei	0	4	4	1
Bass, Anthony	0	2	2	1
Brogdon, Connor	0	1	1	1
Bruihl, Justin	0	4	4	1
Castro, Anthony	0	2	2	1
Fernández, Julian	0	1	1	1
Fleming, Josh	0	2	2	1
Gose, Anthony	0	3	3	1

This data table shows the top 10 highest percent of borderline pitches called for strike. This is not illuminating. The accumulation of lucky calls matters more of rate of lucky calls. I will define a variable `Net_Luck` to refer to this figure: the number of called strikes minus the number of balls. A positive `Net_Luck` value indicates the pitchers is lucky.

player_name	ball	called_strike	total	strike_pct	Net_Luck
Berríos, José	7	27	34	0.7941176	20
Stroman, Marcus	4	23	27	0.8518519	19
Buehler, Walker	7	24	31	0.7741935	17
Cole, Gerrit	8	25	33	0.7575758	17
DeSclafani, Anthony	7	24	31	0.7741935	17
Gray, Josiah	8	25	33	0.7575758	17
Lyles, Jordan	15	30	45	0.6666667	15
Wells, Alexander	13	28	41	0.6829268	15
Cobb, Alex	6	20	26	0.7692308	14
Fedde, Erick	5	19	24	0.7916667	14

This is the ten pitchers with the highest `Net_Luck`

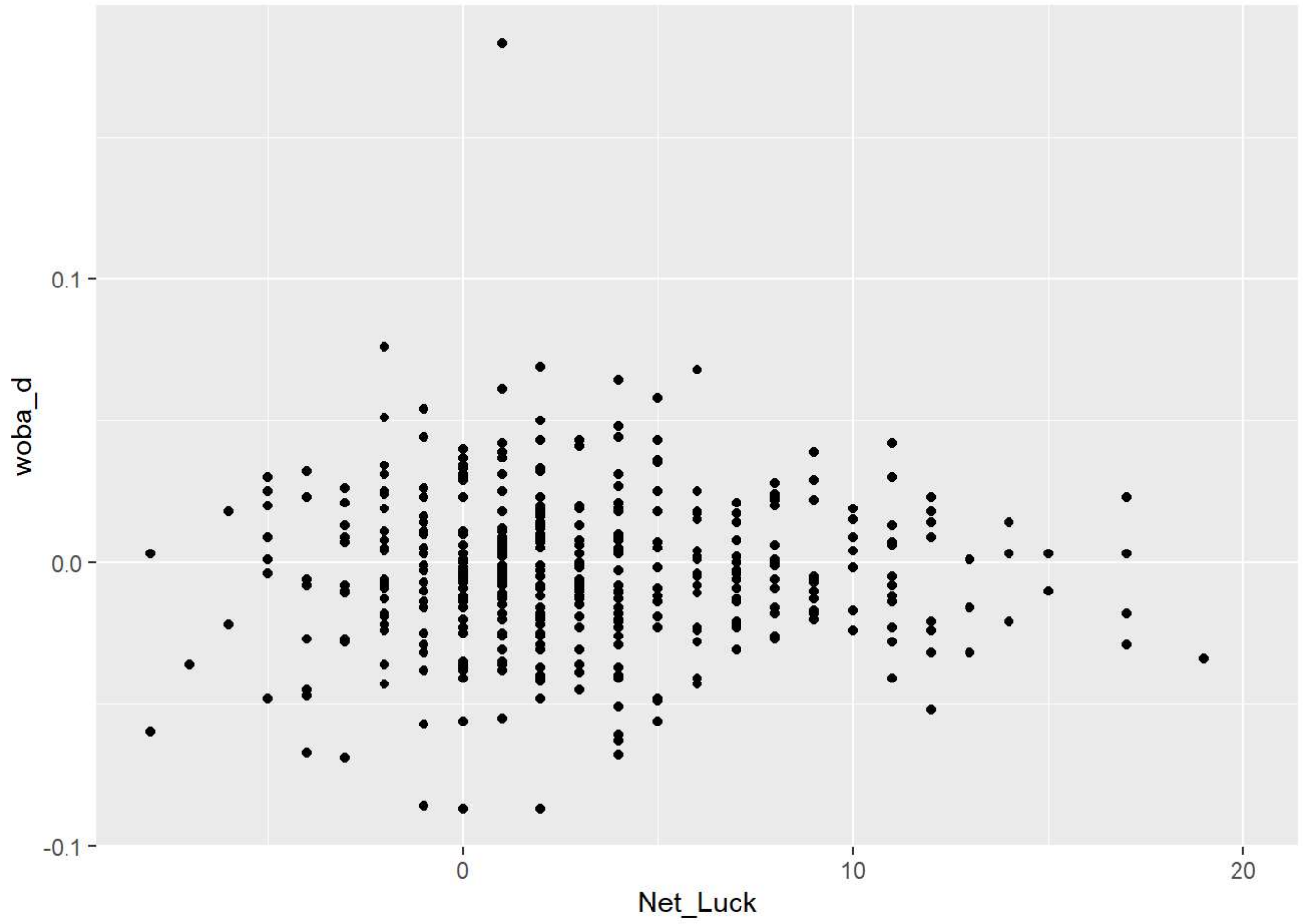
Now let's compare `Net_Luck` to the differences between a pitcher's actual stats and Statcast expected stats.

The real statistics vs. expected statistic difference is calculated like this:

$$\text{Batting Average} - \text{Expected Batting Average} = \text{BA\_Difference}$$

If the result is positive, then the player did better than expected.

We will be looking at wOBA. This graph shows `Net_Luck` on the x-axis and the difference in wOBA vs xwOBA (`woba_d` , for "wOBA differential") on the y-axis. I would expect to see positive `woba_d` values correlated with higher `Net_Luck` values. I have also



There seems to be no such correlation.

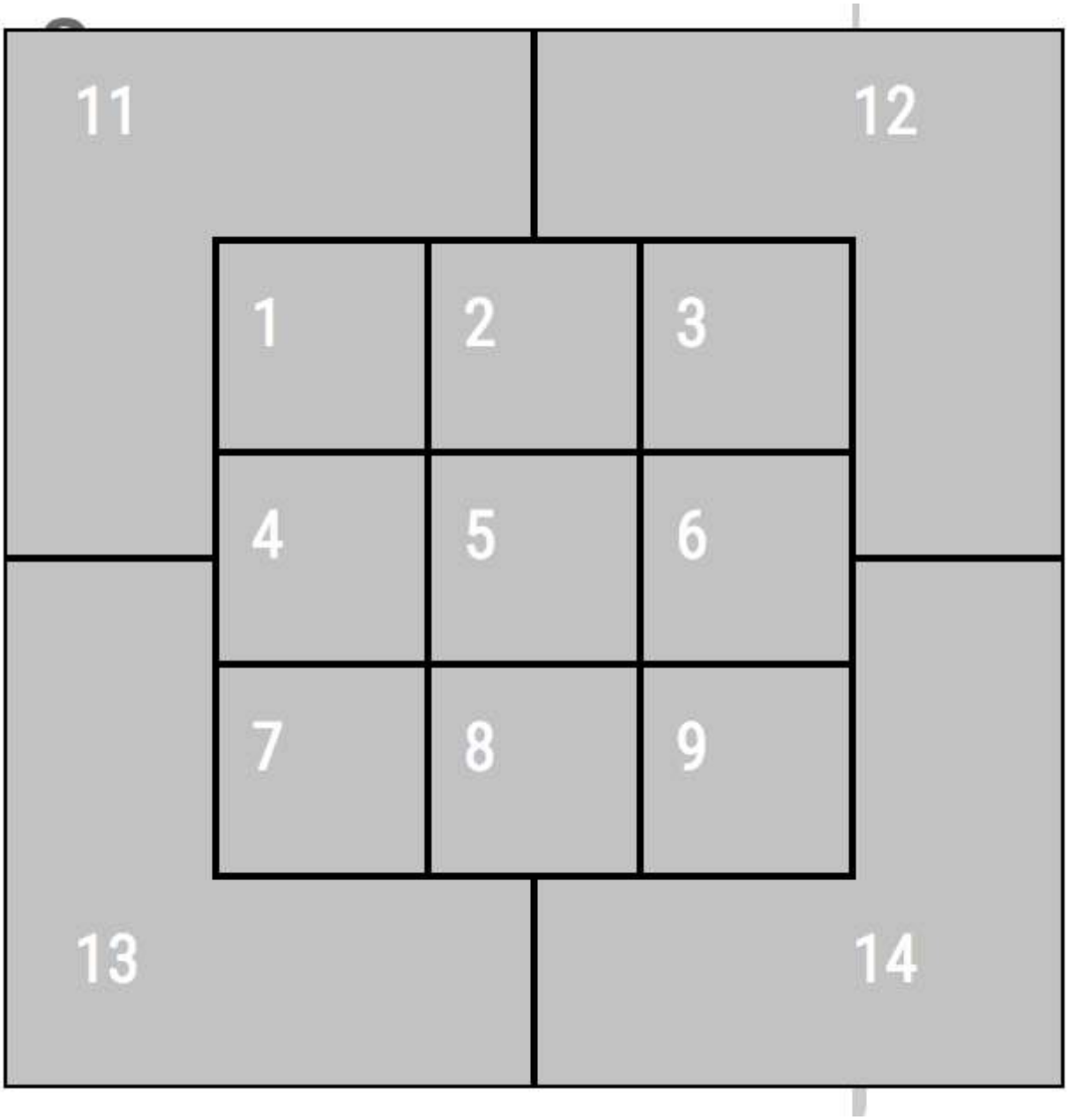
```
cor(Total_Results$woba_d, Total_Results$called_strike, use = "complete.obs")
```

```
## [1] -0.03963996
```

With my definition of luck, there seems to be no correlation between borderline strike calls and out or underpeforming a pitchers expected stats

A Simpler Approach

Let's try a less arbitrary definition of luck: Pitches outside the zone that were called a strike. This is far more straigtforward: Statcast has defined zones 1-9 as strikes, and 11-14 as balls. This time, `SZNet_Luck` - "Strike Zone Net Luck" - will measure luck. The calculation is essentially the same.



player_name	Good_Luck	Bad_Luck	SZNet_Luck
Darvish, Yu	20	4	16
Berríos, José	15	2	13
Gray, Jon	15	3	12
Keuchel, Dallas	14	2	12
Bumgarner, Madison	15	4	11
Senzatela, Antonio	12	1	11
Gibson, Kyle	13	3	10
Eovaldi, Nathan	9	0	9
Hendricks, Kyle	11	2	9

player_name	Good_Luck	Bad_Luck	SZNet_Luck
Ryan, Joe	10	1	9

This list looks a a lot different from out previous `Net_Luck` top 10. Now to compare with `woba_d`

Even with a tighter definition of luck, there doesn't appear to be a correllation between helpful calls and performance relative to expected statistics.

Closing Thoughts

This is puzzling. I set out on this investigation to try to quantify how a bad umpiring could help or hinder a pitcher's season, inspired by the research that shows the importance of gettign ahead in the count. Because the average batter's wOBA is 0.363 after a 1-0 count and 0.270 after a 0-1 count, I expected to see that reflected in this research. (Source: https://blogs.fangraphs.com/the-count-is-king-even-after-accounting-for-batter-skill/ (https://blogs.fangraphs.com/the-count-is-king-even-after-accounting-for-batter-skill/))

It's understandable that there was no correlation when using my definition of borderline. In that definition, pitches that Statcast defined as strikes and were called as strikes, were considered lucky. This would mean that pitchers who consistently peppered the edge of the zone are considered luckier, given that umpires judge correctly.

A possible reason no correlation is seen in the Luck vs wOBA_differential graph using the strike zone to define luck is that there are other, more important luck-based factors. The first statistic that jumps to mind is BABIP. It would be interesting to see if BABIP relates to wOBA differential.