Estimating a Pitcher's Standard Deviation Based on an Assumed Target

Why Use an Estimated Standard Deviation?

Originally, I used the raw MLB data to take the raw standard deviation for each grouping of pitcher, batter, and pitch type. Most of these standard deviations were around 0.6ft both horizontally and vertically. MLB pitchers are very precise, so this seemed unreasonable. The flaw with this approach was that it did not take into account where the pitcher was aiming. There is also no raw data available that indicates how much a pitcher missed their intended target by. This led me to creating an estimate standard deviation based on an assumed target location.

Data Where I Can Assume a Target

The pitches that were easiest to assume a target location were located near any edge of the set strike zone. I filtered the data to pitches located from one ball width inside of the zone to 1 ft outside of the zone, on any edge. The shaded gray area represents where this data falls. There were very few pitches located further than 1 ft outside of the zone, and at that point, I can no longer make a guess as to where they were aiming.

Assumed Target

For the data described above, I assumed pitchers tried to aim 0.11 ft outside of the zone for the edge the pitch was closest to. The radius of a baseball is ~0.12ft, so I chose 0.11 to guarantee that part of the ball would be touching the edge of the strike zone. Even for pitches inside of the zone, I assumed they were intended to be placed ~0.11ft outside.

I made this assumption because of the strike zone analysis portion of this project and the current ABS challenge system in MLB. Batters have a sharp eye, so they will likely challenge more pitches that land more outside of the zone than inside. If pitchers can successfully target ~0.11ft outside of the zone, batters may be more likely to challenge those spots and waste their challenges for the game. Additionally, an umpire is more likely to rule a pitch in these locations a strike rather than a ball, as detailed in my strike zone analysis.