Pitcher VS. Hitter

- Affect of the Count on a Hitter's Average…

Dana Vahey

December 2008

Data Mining Methods

Problem Description

- In Baseball/Softball, there are 12 possible counts that a batter can have while at bat. Some of these counts can be considered in favor of the batter, some the pitcher, and two here are considered neutral. The idea of this project is to look at the different counts of each of the three chosen batters and analyze the data to see if their approach at the plate differs depending on their count. Does the batter swing more for power when they are ahead of the count? On the same note, does the batter swing to make contact and at least put the ball in play when they are behind in the count?

Analysis Technique

- There is not a particular “solution” to this project, rather it is more of an analysis. When looking at all of the outcomes for the three batters during the regular 2006 MLB (Major League Baseball) Season, each individual batter’s separate count is compared with their regular season averages. For example, all three batters have a higher SLG when they have a 2-0 (hitter’s) count compared to their regular season SLG. And at the time, when these three hitters have an 0-2 (pitcher’s) count, their SLG is lower than their regular season SLG.
- By comparing each batter’s count individually with their regular season averages, we can see the difference of how the batter might have changed their approach when at bat.

Other Factors to Consider

- Having people on-base might have an affect on how the hitter approaches the plate
- If the opposing team brings in another pitcher
- The pitcher is a person too; they can have different mind-sets depending on the batter and the batter's count too.
  - For example, if the pitcher is ahead in the count while facing a highly respected hitter like Albert Pujols, they are most likely not going to give him anything too decent to hit (they will try to get Pujols to chase a pitch outside of the [strike] zone instead of allowing him to see a better pitch).
- The beginning attitude that each hitter comes to the plate with every time.
  - Example: Eckstein is a lead-off hitter
  - Example: Pujols hits in the third or fourth spot (a.k.a. the "clean-up spot"),
  - Different players have different approaches…
- These are just a few factors that might have a small impact on the results…
- ((Even though as a player, you should try to have the same solid approach every time)).

FYI - Some Abbreviations

- AVG = Batting Average
- OBP = On-Base Percentage
- SLG = Slugging Percentage
- H = Hit (Single), 2B = Double,
  3B = Triple, and HR = Homerun.
- AB = At Bat
- BB = Base on Balls (Walk)
- HBP = Hit By Pitch
- SF = Sacrifice

Formulas for Calculations

$$AVG = \frac{H}{AB}$$

$$OBP = \frac{H + BB + HBP}{AB + BB + HBP + SF}$$

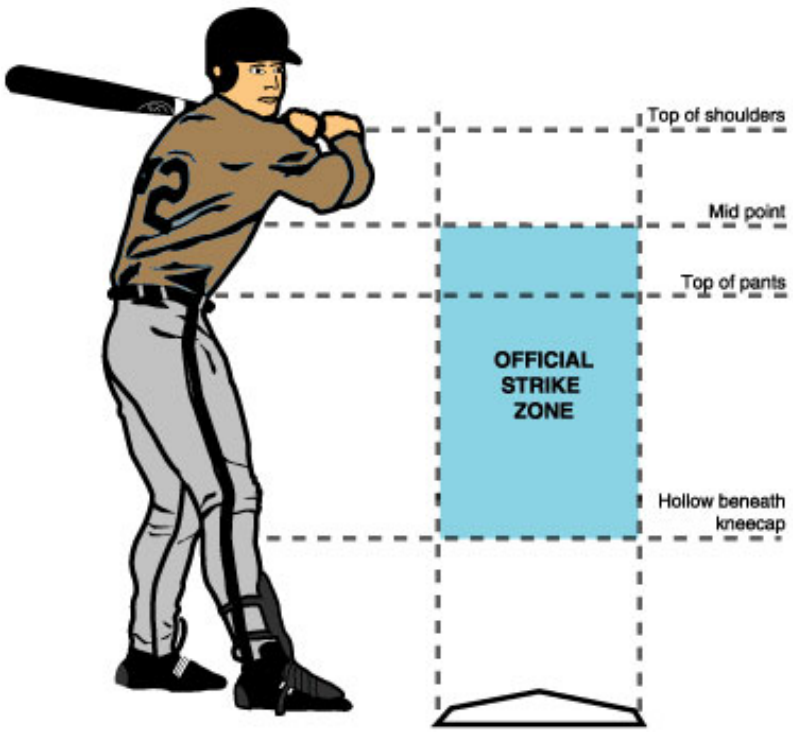
$$SLG = \frac{(1B) + (2 \times 2B) + (3 \times 3B) + (4 \times HR)}{AB}$$

Possible Batting Counts for Pitchers/Hitters

Hitter's Count Neutral Pitcher's Count

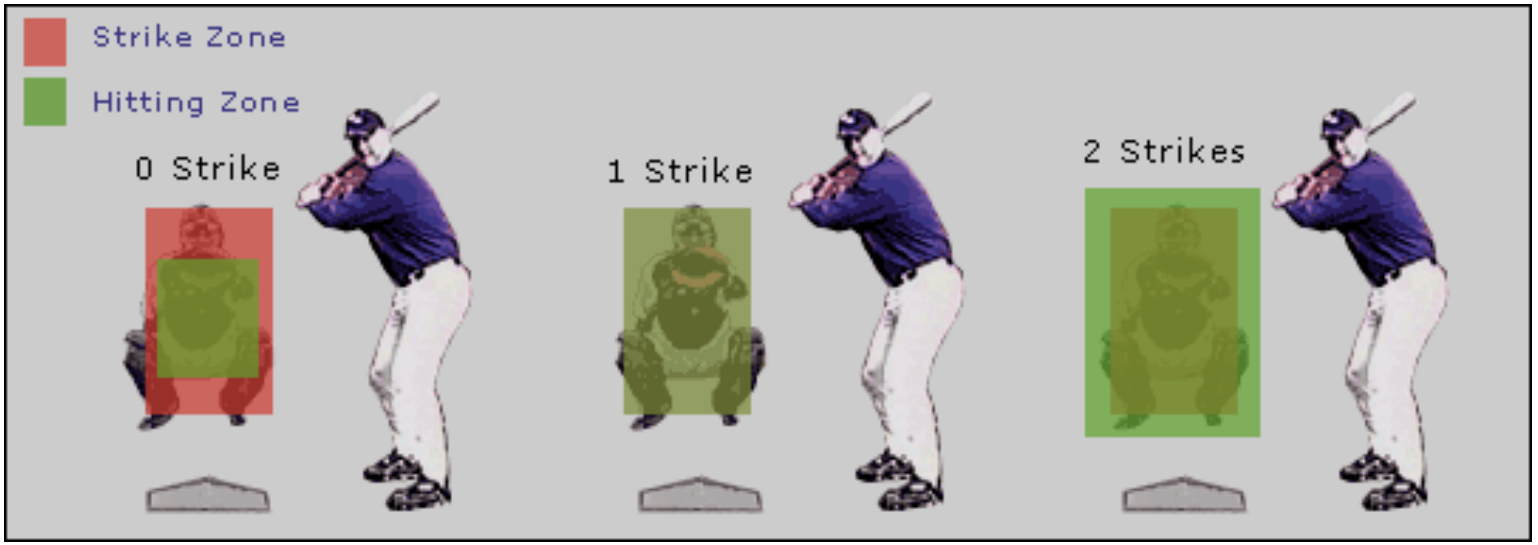
- 0-0 – Hitters' Count
- 0-1 – Pitcher's Count
- 0-2 – Pitcher's Count
- 1-0 – Hitter's Count
- 1-1 - Neutral
- 1-2 – Pitcher's Count
- 2-0 – Hitter's Count
- 2-1 – Hitter's Count
- 2-2 – Pitcher's Count
- 3-0 – Hitter's Count
- 3-1 – Hitter's Count
- 3-2 - Neutral

Official Strike Zone



- Every batter's strike zone varies depending on their height and stance…
- The 2004 Official Rules of Major League Baseball defines it as "that area over home plate the upper limit of which is a horizontal line at the midpoint between the top of the shoulders and the top of the uniform pants (of the batter), and the lower level is a line at the hollow beneath the knee cap."

A visual of how a batter's perception on when to swing might change throughout the count…



2006 Regular Season Batting Statistics

- 3 Players from the St. Louis Cardinals World Series Championship Team -

David Eckstein – ShortStop #22 – 2006 Regular Season Batting Statistics

AVG	OBP	SLG
.292	.350	.344

Albert Pujols – First Base #5 - 2006 Regular Season Batting Statistics

AVG	OBP	SLG
.331	.431	.671

Yadier Molina – Catcher #4 - 2006 Regular Season Batting Statistics

AVG	OBP	SLG
.216	.274	.321

0-0

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Eckstein	.341	.429	.366	.292	.350	.344

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Pujols	.326	.333	.721	.331	.431	.671

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Molina	.347	.382	.569	.216	.274	.321

0-1

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Eckstein	.250	.250	.303	.292	.350	.344

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Pujols	.426	.429	.833	.331	.431	.671

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Molina	.175	.190	.246	.216	.274	.321

0-2

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Eckstein	.204	.218	.204	.292	.350	.344

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Pujols	.167	.167	.444	.331	.431	.671

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Molina	.162	.162	.243	.216	.274	.321

1-0

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Eckstein	.375	.346	.417	.292	.350	.344

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Pujols	.396	.408	.917	.331	.431	.671

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Molina	.273	.273	.424	.216	.274	.321

1-1

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Eckstein	.323	.343	.431	.292	.350	.344

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Pujols	.367	.375	.835	.331	.431	.671

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Molina	.190	.209	.286	.216	.274	.321

1-2

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Eckstein	.244	.277	.267	.292	.350	.344

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Pujols	.319	.319	.507	.331	.431	.671

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Molina	.132	.164	.189	.216	.274	.321

2-0

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Eckstein	.667	.667	.667	.292	.350	.344

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Pujols	.381	.381	1.000	.331	.431	.671

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Molina	.143	.143	.357	.216	.274	.321

2-1

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Eckstein	.479	.479	.563	.292	.350	.344

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Pujols	.327	.327	.596	.331	.431	.671

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Molina	.280	.280	.280	.216	.274	.321

2-2

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Eckstein	.259	.259	.293	.292	.350	.344

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Pujols	.260	.260	.468	.331	.431	.671

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Molina	.149	.143	.170	.216	.274	.321

3-0

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Eckstein	.000	1.000	.000	.292	.350	.344

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Pujols	.000	1.000	.000	.331	.431	.671

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Molina	.000	1.000	.000	.216	.274	.321

3-1

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Eckstein	.400	.800	.600	.292	.350	.344

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Pujols	.364	.736	.591	.331	.431	.671

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Molina	.222	.650	.333	.216	.274	.321

3-2

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Eckstein	.222	.440	.333	.292	.350	.344

Player	BA	OBP	SLG	sBA	sOBP	sSLG
Pujols	.324	.579	.618	.331	.431	.671

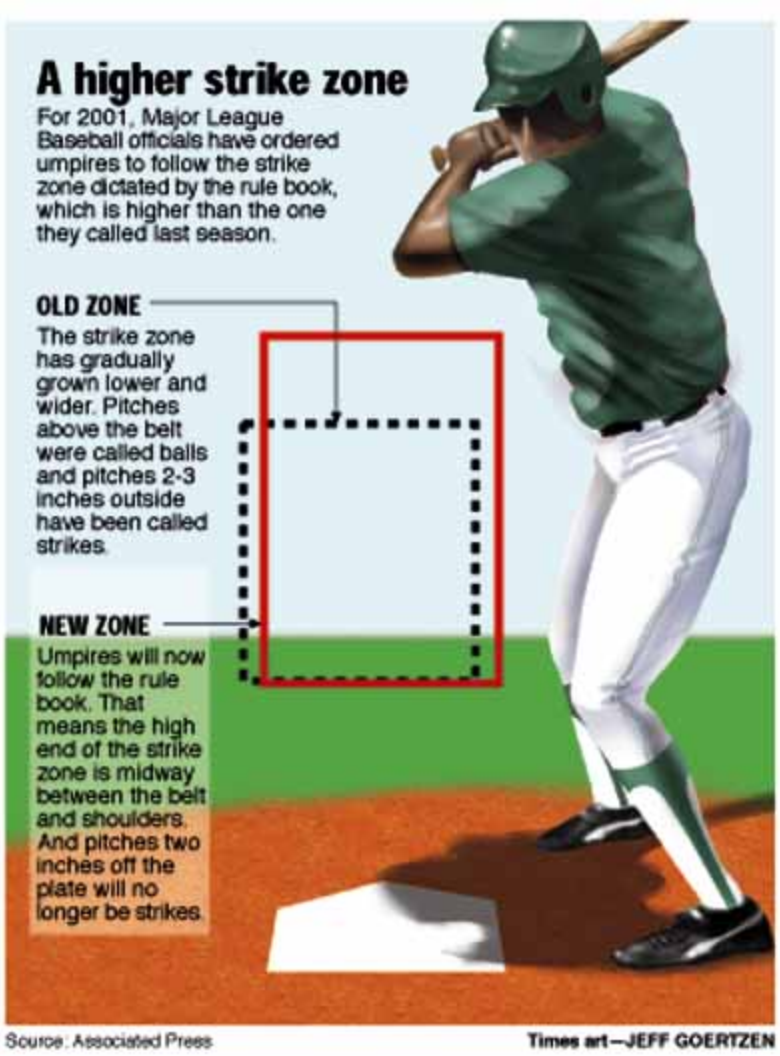
Player	BA	OBP	SLG	sBA	sOBP	sSLG
Molina	.250	.417	.393	.216	.274	.321

Results

- The easiest statistic to measure the affect of a count on the batter's approach at the plate is probably SLG. SLG measures the hitter's power. Therefore, the hypothesis that a hitter would focus **more** on making **contact** and **less** on **power** when they are **behind** in a count would be supported when the SLG for that hitter with that particular count is **lower** than the batter's season SLG. On the flip-side, there would be a strong correlation when looking at when the batter's SLG is higher than their season SLG when they are ahead of the count.
- With a few exceptions, the higher the batter is ahead of the count, the higher their BA, OBP, and SLG is compared to their season averages (especially when looking at SLG). The reverse is noticeable too; the more behind in the count the batter is, the lower their BA, OBP, and SLG is when compared to their season averages. One such exception is when the count is 3-0. A reason for this is it is widely practiced for hitters not to swing when the count is 3-0. Not only might they be walked, but it forces the pitcher to throw another pitch (more work). And besides that, if the pitcher throws a strike, the batter is still ahead in the count with a 3-1 count, and is still likely to see a better pitch to hit than if they were behind…

Issues

- For the most part, the hypothesis is fairly supported. This is mostly visible when looking at the SLG data for each of the three hitters. However, there are a few instances in the data that do not support this theory. These are the averages in white, while the averages in black support the hypothesis (the neutral counts were not altered in color either way).


A higher strike zone
For 2001, Major League Baseball officials have ordered umpires to follow the strike zone dictated by the rule book, which is higher than the one they called last season.
OLD ZONE
The strike zone has gradually grown lower and wider. Pitches above the belt were called balls and pitches 2-3 inches outside have been called strikes.
NEW ZONE
Umpires will now follow the rule book. That means the high end of the strike zone is midway between the belt and shoulders. And pitches two inches off the plate will no longer be strikes.
Source: Associated Press
Times art—JEFF GOERTZEN

References

- www.google.com
- www.wikipedia.org
  - http://en.wikipedia.org/wiki/Slugging_percentage
  - http://en.wikipedia.org/wiki/On_Base_Percentage
- http://sports.espn.go.com/mlb/teams/batting
- www.official-rules.org
- www.sptimes.com
- www.joecannaday.com
- Mr. Aleshunas, J.


OLD STRIKE ZONE.
NEW STRIKE ZONE.