By the Numbers

Volume 17, Number 1 The Newsletter of the SABR Statistical Analysis Committee February, 2007

Review

Academic Research: A Hit-By-Pitch Study

Charlie Pavitt

The author describes a recent academic study investigating hit by pitches. Are a result of a desire for revenge against batters who previously hit a home run off the pitcher? And do pitchers plunk fewer batters when they will be coming to the plate themselves?

This is one of a series of reviews of sabermetric articles published in academic journals. It is part of a project of mine to collect and catalog sabermetric research, and I would appreciate learning of and receiving copies of any studies of which I am unaware. Please visit the Statistical Baseball Research Bibliography at www.udel.edu/communication/pavitt/biblioexplan.htm. Use it for your research, and let me know what is missing.

John Charles Bradbury and Douglas J. Drinen, Crime and Punishment in Major League Baseball: The Case of the Designated Hitter and Hit Batters, Economic Inquiry, January 2007, Vol. 45 No. 1, pages 131-144

In this issue

The question of whether hit batsmen are the result of strategic choice on the part of the pitcher and manager of the opposing team has been examined by several economists over the past decade, and I have previously reviewed one relevant study back in *BTN* Volume 14, Number 2. From this work, we have learned that the presence of the designated hitter since 1973 has substantially increased the number of hit-by-pitches in the American League over both the National League and the AL pre-DH.

Two explanations for this tendency have been proposed: that pitchers are more likely to plunk a batter when they don't bat themselves and so know they won't be the target of retaliation; and that there is less harm giving DHs first base with a HBP than pitchers as they are more likely to get on base in the first place. The retaliation hypothesis has particularly interested these researchers, but all previous studies have relied on aggregate seasonal data, making it difficult to test.

Retrosheet comes to the rescue, and Bradbury and Drinen use its data to determine whether a hit-by-pitch in one half inning increases the odds of retaliation in the next. According to two analyses (one for 1969 combined with 1972 through 1974, the other for 1989 through 1992), it does, as does a home run by the previous batter in the more recent data set; both of these findings support the retaliation hypothesis.

However, higher OPS was positively associated with HBP whereas pitchers were less likely to be plunked than everyone else; both of these results suggest the "less harm" hypothesis. In addition, large score differentials increase HBP, likely because there is less harm when such a differential leaves less doubt concerning which team will probably win the game. Finally, wilder pitchers are, not surprisingly, more likely to hit batters.

Bradbury and Drinen also noted that HBP exploded during the 1990s, particularly in the case of the National League, whose numbers came to approximate that of the American despite the absence of the DH. The authors believe this to be a perverse result of the rule change authorizing umpires to warn both teams not to retaliate, as it lowers the chance that pitchers will be plunked, thus (according to the retaliation hypothesis) leading them to feel free to throw at hitters.

In summary, Bradbury and Drinen have offered us the best work on hit by pitches of which I am aware..

Charlie Pavitt, chazzq@udel.edu ♦

Get Your Own Copy

If you're not a member of the Statistical Analysis Committee, you're probably reading a friend's copy of this issue of BTN, or perhaps you paid for a copy through the SABR office.

If that's the case, you might want to consider joining the Committee, which will get you an automatic subscription to BTN. There are no extra charges (besides the regular SABR membership fee) or obligations – just an interest in the statistical analysis of baseball.

The easiest way to join the committee is to visit http://members.sabr.org, click on "my SABR," then "committees and regionals," then "add new" committee. Add the Statistical Analysis Committee, and you're done. You will be informed when new issues are available for downloading from the internet.

If you would like more information, send an e-mail (preferably with your snail mail address for our records) to Neal Traven, at beisbol@alumni.pitt.edu. If you don't have internet access, we will send you BTN by mail; write to Neal at 4317 Dayton Ave. N. #201, Seattle, WA, 98103-7154.

Submissions

Phil Birnbaum, Editor

Submissions to *By the Numbers* are, of course, encouraged. Articles should be concise (though not necessarily short), and pertain to statistical analysis of baseball. Letters to the Editor, original research, opinions, summaries of existing research, criticism, and reviews of other work are all welcome.

Articles should be submitted in electronic form, either by e-mail or on CD. I can read most word processor formats. If you send charts, please send them in word processor form rather than in spreadsheet. Unless you specify otherwise, I may send your work to others for comment (i.e., informal peer review).

If your submission discusses a previous BTN article, the author of that article may be asked to reply briefly in the same issue in which your letter or article appears.

I usually edit for spelling and grammar. If you can (and I understand it isn't always possible), try to format your article roughly the same way BTN does.

I will acknowledge all articles upon receipt, and will try, within a reasonable time, to let you know if your submission is accepted.

Send submissions to:
Phil Birnbaum
88 Westpointe Cres., Nepean, ON, Canada, K2G 5Y8
birnbaum@sympatico.ca

Is Walk the Opposite of Strikeout?

Russell Carleton

Commentators sometimes treat walks and strikeouts as if they were the opposite of each other – players with lots of walks have "good plate discipline," while players with lots of strikeouts do not. Here, the author explains why this line of thought is incorrect, and comes up with new statistical measures of the batter's strike-zone judgement.

Commonly, baseball researchers have used the strikeout-to-walk ratio as a measure of a hitter's plate discipline or "batting eye." The logic behind the ratio seems relatively obvious: the two outcomes of a plate appearance that don't involve the ball being put into play are the walk and the strikeout. As both events are made up by either good or bad command of the strike zone, it would seem that they are the opposite sides of the same coin.

There's a problem: a batter can strike out more than one way. More to the point, a batter can *avoid* striking out in more than one way. Consider a few extreme examples. Freddy Freeswinger swings at every pitch he sees. Freddy will never take a base on balls, and he may very well strike out. If he avoids a strikeout, it will be because he hit the ball in play. Peter Patient, on the other hand, takes every pitch that he is thrown. He might be called out on strikes, but he might also walk if the pitcher is wild. Of course, no such players exist, but players do vary in how often they swing. In 2006, Scott Hatteberg swung the least often (34% of the pitches thrown to him) among hitters with more than 100 PA. Delmon Young swing the most often (68% of the time, a full 7 points higher than the next closest hitter, A. J. Pierzynski.)

For a free-swinging hitter, the important thing to consider is how often he makes contact when he swings. Because he swings a great deal, he's not likely to walk, but if he makes contact with pitches (either putting them in play or fouling them off), he's less likely to strike out. In this case, his walk total (in Freddy's theoretical case, zero) will be largely determined by his free-swinging ways, but his strike out total will be determined by his ability to make contact. For a patient hitter, the formula is a little different. A patient hitter who simply takes everything is at the mercy of the control of the pitcher. However, a patient hitter may be able to take several pitches out of the strike zone (and probably draw a few walks), but have the good discernment to know when a good hittable pitch comes over the plate and to politely deposit it in the left field seats. So, any player's strikeout and walk totals (and thus his K/BB ratio) will be determined by the interaction of a few different factors: how well he understands the strike zone, how often he swings, and whether or not he makes contact with the pitch.

It is my contention that a simple strikeout to walk ratio does not accurately describe a hitter's full range of abilities in terms of plate discipline, and that new measures, plural, are needed. I also contend that the abilities associated with the likelihood that a batter will walk and the likelihood that a batter will strike out are manifestations of *different* abilities, or at least that the recipe of a walk is different from that of a strikeout, even if they are made from the same ingredients. While I agree with the thought that a disciplined hitter is one who can avoid striking out, I do not believe that it follows that a player who draws a great number of walks is necessarily "disciplined." A disciplined hitter may instead be one who waits for his pitch and then hits it into play. A hitter who draws a large number of walks may simply be taking too many pitches. I propose that what has previously been called plate discipline can be better understood as a combination of being able to properly discern whether pitches are in the strike zone (and generate the proper response) and creating a proper balance between swinging too much and swinging too little. The new metrics that I propose are based on the concept of signal detection theory, a concept originally developed for medical epidemiological studies which has been used extensively in the fields of cognitive psychology and engineering.

For those unfamiliar with signal detection theory, the idea is relatively simple. Suppose that there is a nasty virus going around, and a researcher wants to develop a test that will detect it. The test he develops returns a result of either "yes" (the virus or signal is present) or "no" (the virus or signal is not present), but it's not yet established that those answers are accurate. The proper research method is to gather together a group of individuals who have already been diagnosed as having the virus by other means and a group already known not to have the virus and to run the test on them. If the test properly classifies all those known to have the virus as "yes" and those without as "no", it is a perfect test. Sadly, no test is perfect. There are always errors, but it's important to know what kind of errors are made as they can be of two *different* types. Suppose that the test reports that someone who is actually known to be healthy has the virus. This is what's known as a "false positive" or Type I error. The opposite case, where a person known to have the virus is reported as being healthy would be a "false negative" or Type II error.

	Virus is truly present	Virus is not truly present
Test says virus is present	Correct	Type I Error
Test says virus is not present	Type II Error	Correct Rejection

A test that commits fewer errors is known as a sensitive test. But, if the test makes one type of errors more than the other, this also tells us something about the test. If the test produces a great deal of false positives, but relatively few false negatives, it means that the test has a bias toward saying yes. In the extreme case, the test will identify everyone, sick and well, as having the virus. While the researcher has created a test that has identified all of the sick people, it's a useless test, as it simply says "yes" in all situations. The same scenario can happen with false negatives, in which case the test may have a bias toward saying no. Therefore, the ideal test has a high sensitivity and a response bias that is neither too far in the direction of saying yes or no. Those with an interest in the gory details of how these statistics are calculated may consult the appendix. Statistically, one would hope that the sensitivity of a measure would be as high as possible and the measure of response bias (also called the criterion) would be near 1. A response bias above 1 indicates that a test is more likely to produce a "yes" response (at the cost of additional Type I errors), while a response bias below 1 indicate a greater bias toward a "no response" (at the cost of additional Type II errors).

Returning to baseball, a batter faces an analogous situation. A pitch is thrown and it will either be in the strike zone or out of the strike zone. Either way, the batter must decide whether or not he will swing. If the pitch is in the strike zone, the proper response from the batter is to swing at the pitch. If he lets it go by, he's guaranteed to take a strike, but if he swings, he may hit it. However, if the pitch is not in the strike zone, the best response from the batter is not to swing. If he doesn't swing, it will be a ball. If he does swing, it will likely be a swinging strike (although he might end up hitting it anyway). In this case, whether or not the ball is in the strike zone is analogous to whether the virus is present. The batter's "eye" is the test; that is whether he can properly determine the difference between a strike and a ball and then make the appropriate response for the situation, to swing or not to swing.

There are a few difficulties with this framework: One might argue that it is not always advantageous, either in baseball or in medicine to have a response bias of exactly 1.00. Indeed, many tests have a response bias set above 1.00 (although not too much above), specifically because false positives (telling someone that they have the disease when they do not), can be double checked by other tests, but false negatives (telling someone that they are healthy when they actually have the disease), can lead to a patient not receiving treatment. In this case, the cost of a Type I error and the cost of a Type II error are not the same.

In baseball as well, some hitters change their response bias situationally, for example taking a 3-0 pitch. If a batter completely forgoes swinging, there are only two possible outcomes, a called strike (and a 3-1 count) or a ball to complete the walk. That hanging curveball will just have to go by. The idea there is that the cost of a strike is relatively minimal, but the benefits of a ball are great (first base for free). On the other end of the spectrum, a batter might be more likely to swing on a two-strike count to "protect" the strike zone. There are also times when a batter will knowing let a perfectly good strike go by because he is waiting for "his" pitch, preferably one that he can drive. As the song reminds us, a batter does get 1-2-3 strikes before he is out. Taking a pitch for a strike would count as an error in this framework, although it might be more advantageous in the long run. The framework is not perfect, but overall, hitters who take fewer strikes will come out ahead.

These situations, while they do happen, are relatively rare when compared to the couple of thousand pitches that a batter faces during a season. In general, the cost of the two "errors" in this framework (a swinging strike vs. a called strike) is the same. After all, the umpire clicks the same button either way. In most situations, the proper response bias is 1.00. At a response bias of 1.00, a player is effectively minimizing how many errors (strikes) he will make in managing the count. If he swings too much (and his response bias goes above 1.00), he will mathematically put more balls in play, but he will purchase those extra balls in play with an even greater number of swinging strikes. The further above 1.00 he goes, the more swinging strikes he will have to give up to hit an extra ball. The same pattern holds below 1.00. A hitter below 1.00 will take extra balls, but will purchase them at the cost of a greater number of called strikes. On the whole, a hitter below 1.00 would gain more by swinging more than he would lose (in terms of balls and strikes), and a hitter above 1.00 would gain more by swinging less. A hitter whose response bias is exactly 1.00 will have the fewest number of strikes called against him, *given his abilities*.

Thanks to Retrosheet, those of us in the general public have access to pitch-by-pitch data for most games played over the past few decades, complete with balls and strikes (including whether the strike was called, swinging, or a foul ball). For the purposes of the initial validation, I gathered together the game data for the entirety of Major League Baseball for the years 1993-1998. After some data crunching, I was able to count the number of occurrences for each type of pitch event (i.e., ball, swinging strike, etc.) for each batter during the course of each individual year. This produced a database of 5065 player-seasons, although the analyses presented in this paper only include those batters who had more than 100 plate appearances in the season in question. This limited the data set to 2426 player-seasons, an average of 404 batters per season, the reduction largely owing to National League pitchers batting, as well as players who received "cup of coffee" call ups.

Within the signal detection framework, I classified a swinging strike as a "false positive" (the batter responded by swinging when no response would have been the better course of action), while a ball in play was a correct response (the batter responded by swinging which was the proper thing to do). A called strike was a "false negative" (the batter did not respond when he should have), and a ball as a correct non-response. Foul balls proved somewhat more difficult to classify. Clearly, the act involves swinging, so the batter has produced a response, but when the batter has either zero or one strike in the count against him, a foul ball is equivalent to a swinging strike, while a two-strike foul ball does not affect the count. After trying several variations, I settled on classifying zero and one-strike foul balls as swinging strikes, while two-strike foul balls were "correct responses." The framework isn't perfect in that a batter might swing at a pitch that is in the strike zone and still miss (fooled by a change-up, guessed the wrong part of the plate). He might hit a ball out of the strike zone into play.

Taking a pitch for a strike might be advantageous in some situations. This framework is centered around the batter's decision to swing and rather than his discernment of the strike zone proper (actual pitch *location* data was not immediately available in a useable format), it looks at the *results* of his swinging (or not swinging) as a proxy for pitch location. After all, it's the results that actually matter. In an ideal world, however, I would have preferred to see whether batters might be swinging at pitches in the strike zone, but missing them.

Using the signal detection framework, I calculated each player's sensitivity and response bias statistics for each year. The sensitivity statistic in this case is a reflection of how good a batter is at judging between the pitches at which he should and shouldn't swing. The response bias statistic is a statistic of how likely the batter is to swing. Again, my methodology is in the appendix.

I also calculated each player's strike out and non-intentional walk rate, per plate appearance. As the variables are binomial in nature, I took the natural log of the odds ratio to normalize the distribution.

Results

The new statistics proved remarkably stable over time, with an intra-class correlation coefficient of .720 for the sensitivity measure and .807 for the response bias measure[1]. For those unfamiliar with this method, it looks at players over time and finds how much of the variance in the measurement (in this case, our newly created sensitivity and response bias measures) is attributable to the players themselves. In this case, approximately 50% (.720 * .720) of the variance in sensitivity is attributable to the players over the six years in the database; 65% of the variance in response bias is similarly attributable to the players themselves from year to year. The sensitivity measure correlated with K/BB ratio at -.497, while the response bias measure correlated with K/BB at .458. Clearly, the two new variables are related to K/BB, but not as closely as one might think if the two were measuring the same construct.

Sensitivity correlated with age at .216 and response bias correlated with age at -.071 suggesting a weak tendency for players to become more sensitive as they age and a very weak relationship between age and willingness to swing. Interestingly, neither variable was associated much with what happened to the ball when the ball was put into play. Correlations between the number of ground balls, pop flies, line drives, and fly balls per ball in play and the two variables proved to be very weak, the strongest being -.111. K/BB ratio showed similar properties. The one area where K/BB ratio had a clear advantage was in correlating more clearly with OPS (-.303 vs. .169 for sensitivity and -.094 for response bias), although as strikeouts and walks appear in the formulae for both K/BB and OPS, one could make the case that the measures are confounded.

I generated a correlation matrix between the newly created variables (sensitivity and response bias) and the strikeout and walk rates. All correlations were significant at the .05 level, although this is mainly due to the large sample size present.

	Sensitivity	Response Bias	Strikeout Rate	Walk Rate
Sensitivity	1.000	-.152	-.839	.107
Response Bias		1.000	.229	-.594
Strikeout Rate			1.000	.064
Walk Rate				1.000

Several interesting findings emerge. First, on a more technical note, the correlation between the two created statistics (-.152) suggests that the two are relatively independent of one another. However, even more *un*-correlated are strikeout rate and walk rate (.064). If walk truly was the opposite of strikeout, then one might expect that the correlation between the two would be large in magnitude and negative in valence. However, the correlation that exists between the two is actually slightly positive. A hitter who walks more often is actually slightly *more* likely to strike out. The small value of the correlation, however, suggests that a player's likelihood of walking and striking out are largely unrelated.

There also appears to be a split as to how the new statistics correlate with strikeouts and walks. Sensitivity is clearly more associated with strikeouts (-.839), in that more sensitive hitters are less likely to strike out, than with walks (.107). Response bias, on the other hand, is more associated with walks (-.594), in that players who are less likely to swing are more likely to walk, than with strikeouts (.229). The evidence suggests that walk and strike out are two very different concepts.

2006 Players

The top 20 and bottom 10 hitters for each of these statistics in 2006 are presented in Tables 1 and 2. Again, only players with more than 100 PA were eligible.

The players on the top 20 list for sensitivity are a pretty good bunch, many of them known for having good plate discipline, although they aren't uniformly world-beaters. Neifi Perez and his .243/.260/.316 line from 2006 make an appearance on the list. The point to be made here is that while Perez managed the strike zone well and put a lot of balls into play (8 BB and 25 K in 316 PA), what the ball did when he actually put the ball into play was not that impressive. Plate discipline is a good tool to have, but it does not guarantee success with the bat. The players in the bottom 10 of the sensitivity list are mostly spare parts. Statistically, those on the bottom of this list are considered to be functioning near chance in their selection of pitches at the plate. In other words, they are guessing. The top 20 list for response bias is interesting in how little deviation there is from the optimal number (1.00). Indeed, of the 400+ hitters with at least 100 PA in 2006, more than a third were between the range of .95 and 1.05. This suggests that major league hitters are, by and large, given to swing at about the right rate for their abilities.

Table 1 – Top 20 players in Sensitivity and Response Bias

Sensitivity Top 20		Response Bias Top 20 (Ranked by absolute deviation from 1)	
1. Vladimir Guerrero	1.12	1. Rondell White	1.000
2. Moises Alou	1.04	2. Choo Freeman	1.000
3. Bill Mueller	1.01	3. Chris Duffy	.999
4. Kenny Lofton	1.01	4. Brad Hawpe	1.001
5. Brian Giles	.99	5. Lance Berkman	.999
6. Nomar Garciaparra	.99	6. Aramis Ramirez	.998
7. Scott Hatteberg	.97	7. Hector Luna	.998
8. Luis Gonzalez	.96	8. Emil Brown	1.002
9. Luke Scott	.93	9. Ron Belliard	.998
10. Brian McCann	.93	10. Gerald Laird	1.002
11. Johnny Estrada	.93	11. Wilson Betemit	1.002
12. A.J. Pierzynski	.93	12. Brian Schneider	1.002
13. Luis Castillo	.93	13. Lastings Milledge	1.002
14. Neifi Perez	.93	14. Cory Sullivan	1.002
15. Todd Helton	.92	15. Rob Mackowiak	.997
16. Jeff Francouer	.92	16. So Taguchi	1.003
17. Mark Kotsay	.90	17. Gary Bennett	.997
18. Omar Vizquel	.90	18. Danny Ardoin	.997
19. Barry Bonds	.90	19. Eduardo Perez	.996
20. Hideki Matsui	.90	20. J.D. Closser	1.005

Table 2 – Bottom 10 players in Sensitivity and Response Bias

Sensitivity Bottom 10		Response Bias Bottom 10 (Ranked by absolute deviation from 1)	
1. Antonio Perez	.22	1. Vladimir Guerrero	1.671
2. Kelly Shoppach	.28	2. Jeff Francoeur	1.493
3. Sal Fasano	.28	3. Delmon Young	1.449
4. Edgardo Alfonso	.30	4. A.J. Pierzynski	1.445
5. Dallas McPherson	.31	5. Johnny Estrada	1.376
6. Rene Rivera	.35	6. Scott Hatteberg	.656
7. Chris Snelling	.36	7. Luis Castillo	.684
8. Matt Kemp	.37	8. Brian Giles	.718
9. Brad Wilkerson	.37	9. Vinny Castilla	1.272
10. Danny Ardoin	.37	10. Toby Hall	1.267

Interestingly, seven of the ten of the worst in response appear in the top twenty for sensitivity. Guerrero, Francouer, Pierzynski, and Estrada all err on the side of swinging too much, and Hatteberg, Castillo, and Giles all swing too little. Guerrero being the most extreme case makes for the best example, as he is notorious for swinging at anything thrown in the remote vicinity of the plate. Guerrero's "secret" is that when he swings, he generally makes contact. Judging by the results, he must be doing something right, but his response bias means that he is sacrificing a lot of swinging strikes for a just few more balls in play. According to this metric, Guerrero could actually be a better hitter (or at least maximize his balls and strikes better) if he swung less. Guerrero and others who are high sensitivity, high response bias types tend to have OBPs not all that far removed from their AVG, because

[1] For those interested, I used an AR1 (auto-regressive) covariance matrix to calculate these coefficients.

they simply don't stop swinging long enough to walk all that often. On the other hand, Scott Hatteberg's OBP in 2006 was a full 100 points above his AVG (.289 to .389). He was very selective and didn't swing much, leading to a lot of walks. So, to get a full profile of a player, both statistics must be considered in tandem.

Further investigations on the anatomy of walks and strikeouts

To further explore the anatomy of the walk and the strikeout, I decided to look at how other factors that the batter brings to the plate affect the likelihood that he will strike out. I calculated several additional statistics that might logically affect a hitter's control of the strike zone. To that end, I calculated the percentage of pitches in which the batter made the "correct" decision[2], consistent with the definitions used previously in the signal detection model. I also calculated the percentage of pitches at which the batter swung, and the percentage of time that the batter made contact when swinging (even if a foul ball). Additionally, I calculated the average number of pitches per plate appearance seen by the batter (perhaps players who see more pitches in an at-bat have a better chance of figuring out the pitcher and/or umpire's handling of the strike zone?) and the average number of two-strike foul balls a hitter has per plate appearance in which the count reached two strikes (the ability to "fight off" pitches?). Finally, I input a player's age for the year (as of July 1) and his height in inches (taller players have a larger strike zone). For the binomial outcome variables (swing, contact, and good decision percentages), I again took the natural log of the odds ratio to normalize the distribution. For the following analyses, I did not use the two variables, sensitivity and response bias, created above. The three percentages calculated (swing, contact, and good decision) already roughly approximate this information and preliminary testing showed that sensitivity and response bias were superfluous to the results.

Given the large amount of inter-correlation among the variables used as predictors, I used a stepwise regression model to discern which of the above predictors was most important in predicting strikeout and/or walk rates. A stepwise regression looks at all the variables entered as predictors and determines which is the most closely related to the dependent variable. It then runs a regression with that variable alone and saves the residual values, thus eliminating the common variance that the regressed variable shares with each of the other predictors and the dependent variable. The process then repeats until no further significant predictors are found.

I first ran the equation for strikeout rate. The results were as follows:

Variable Entered	Change in R-squared	Standardized Coefficient in Final Model
1. Contact rate	.738	-0.666
2. Good decision rate	.059	-0.471
3. Swing percentage	.059	-0.320
4. Two-strike fouls	.014	0.144
5. Pitches per PA	.001	0.034
6. Age	< .001	-0.018
7. Height	< .001	-0.018

The final R-squared value was .871.

These results have some interesting nuggets in them. Clearly, the measure most closely correlated with the likelihood of striking out is the rate at which the player makes contact with the pitch when he swings, accounting for the first 73.8% of the variance. As expected, a player who makes contact more often strikes out less. Surprisingly, making good decisions on pitches entered into the regression secondly and explained only an additional 5.9% of the variance in strikeout rates, suggesting that developing a good "eye" is not as important as simply learning to make contact with the ball when swinging at it. Swing percentage also entered into the regression, explaining an additional 5.9% of the variance, such that hitters who swung more were *less* likely to strike out. Fouling pitches off with two strikes entered into the regression equation, explaining an additional 1.4% of the variance, but in such a way that fouling more pitches off with two strikes led to more strike outs. While fouling off a two-strike pitch may give the hitter a second (or third or fourth) chance at doing something productive with his plate appearance, it also gives the pitcher a second chance to strike him out. These data suggest that the way to avoid strikeouts is not to teach batters to walk, but instead to teach them to put the ball into play. The additional three variables entered into the regression, but only explaining minimal amounts of additional variance. The most surprising finding was that taller players, despite their larger strike zone, strike out less. Indeed, a taller player is more likely to have longer arms and be more able to reach balls, perhaps even off the outside part of the plate.

The analogous regression for walk rate follows:

[2] The formula here is (non-intentional balls + in play + two strike fouls) / (total pitches – intentional balls)

Variable Entered	Change in R-squared	Standardized Coefficient in Final Model
1. Swing percentage	.535	-0.824
2. Contact rate	.053	-0.633
3. Good decision percentage	.072	0.398
4. Two-strike fouls	.025	0.193
5. Age	.002	0.036
6. Pitches per PA	< .001	0.038

The final R-squared value was .687.

Here, the first variable that enters into the regression equation is the percentage of pitches at which the batter swings, explaining more than half the variance. Like with strikeouts, batters who swing more walk less. The careful reader will note that with the exception of "good decision percentage", the factors that enter into this regression all do so in the same direction that they do with strikeouts. If one interprets good decision percentage as a measure of a batter's "eye", then it could be considered to make the difference between walking and striking out. However, the additional value that it adds to the equation's R-squared value in both walks and strikeouts is overwhelmed by other variables.

Discussion

Overall, the results suggest that walk rates and strikeout rates are relatively uncorrelated, and that walks and strikeouts represent two different skill sets on the part of batters. What is commonly termed the batter's "eye" or "plate discipline" is a relatively small piece of the makeup of that set of skills. If anything, it appears that walk and strikeout are more similar than different and that their opposite is actually putting the ball into play. Therefore, it appears that K/BB ratio is logically a poor measure of plate discipline. This is not to say that it is not a useful statistic, just that it should not be understood to be something it is not.

I do not doubt the existence of plate discipline as a skill, nor would I suggest that it be ignored by managers and general managers. To the contrary, my problem is that plate discipline is too narrowly understood, even within the sabermetric community, as simply something that produces walks and strikeouts. Indeed, a hitter with good plate discipline *should* take a hittable pitch in the strike zone and put it in play. Nearly nothing good can come from a strikeout, save the occasional wild pitch on strike three, but a ball in play might actually become a hit.

I propose that my sensitivity and response bias measures, in tandem, are a more proper metric for plate discipline. Sensitivity more accurately represents how well a player *understands* the strike zone independent of how often he swings, while response bias represents whether or not he maximizes those abilities and whether he should be counseled to swing more or less. The problem is that these measures of plate discipline appear to only be mildly correlated to OPS, which is generally considered to be a good offensive barometer. This means that a selective hitter is not necessarily going to be a good hitter once the ball leaves the bat.

The final question would be whether plate discipline, as I (or anyone else for that matter) have defined it is really all that big a deal. A player who produces a great deal of offensive output, whatever metric is used, but does so in an "undisciplined" manner is still a player who produces a lot of results. The answers to those questions are not immediately available. However, if plate discipline does not correlate with any offensive metric, then this is a challenge to the "old school" maxim that players should learn plate discipline to become good hitters.

Ideally, further research could look at how these metrics are affected by situational factors (are players more or less sensitive in two-strike counts?), and how they are related to other statistics (runs created, OPS, etc.) Also, analogous research might be undertaken for pitchers to see whether their performances can be similarly modeled.

Appendix

And now the actual math behind calculating the sensitivity and response bias statistics. After counting and classifying the pitches faced by a batter over the course of a season: Correct swing, Type I error (swinging strikes), Type II error (called strikes), and correct rejections (balls). Calculate the "correct swing rate" (which shows how often a batter correctly responded to a strike that was present) by the formula (correct swing) / (correct swing + called strikes). Calculate the "false alarm rate" by the formula (swinging strikes) / (swinging strikes + balls). Convert the correct swing rate and the false alarm rate to z-scores. This can be done by using the Inverse Distribution Function of your statistical program (set the mean to zero, and standard deviation to one: the z-distribution). If you do not have this capability, there are web-based calculators that will do it for you. The sensitivity measure (called d' among statisticians) is the z-score for the false alarm rate minus the z-score for the hit rate.

The formula for response bias is slightly more difficult. First, calculate the phi statistic for the hit-rate z-score. This can be calculated

$$\frac{e^{-z\_hits*z\_hits}}{\sqrt{2\pi}}$$

The phi statistic for the false alarm z-score can be calculated analogously.   The response bias statistic is then given by the formula false_alarm_phi / hit_rate_phi.

Russell Carleton, RCARLETO@depaul.edu ♦

Informal Peer Review

The following committee members have volunteered to be contacted by other members for informal peer review of articles.

Please contact any of our volunteers on an as-needed basis - that is, if you want someone to look over your manuscript in advance, these people are willing. Of course, I'll be doing a bit of that too, but, as much as I'd like to, I don't have time to contact every contributor with detailed comments on their work. (I will get back to you on more serious issues, like if I don't understand part of your method or results.)

If you'd like to be added to the list, send your name, e-mail address, and areas of expertise (don't worry if you don't have any - I certainly don't), and you'll see your name in print next issue.

Expertise in "Statistics" below means "real" statistics, as opposed to baseball statistics - confidence intervals, testing, sampling, and so on.

Member	E-mail	Expertise
Shelly Appleton	slappleton@sbcglobal.net	Statistics
Ben Baumer	bbaumer@nymets.com	Statistics
Jim Box	jim.box@duke.edu	Statistics
Keith Carlson	kcsqrd@charter.net	General
Dan Evans	devans@seattlemariners.com	General
Rob Fabrizzio	rfabrizzio@bigfoot.com	Statistics
Larry Grasso	l.grasso@juno.com	Statistics
Tom Hanrahan	Han60Man@aol.com	Statistics
John Heer	jheer@walterhav.com	Proofreading
Dan Heisman	danheisman@comcast.net	General
Bill Johnson	firebee02@hotmail.com	Statistics
Mark E. Johnson	maejohns@yahoo.com	General
David Kaplan	dkaplan@UDel.Edu	Statistics (regression)
Keith Karcher	karcherk@earthlink.net	Statistics
Chris Leach	chrisleach@yahoo.com	General
Chris Long	clong@padres.com	Statistics
John Matthew IV	john.matthew@rogers.com	Apostrophes
Nicholas Miceli	nsmiceli@yahoo.com	Statistics
John Stryker	john.stryker@gmail.com	General
Tom Thress	TomThress@aol.com	Statistics (regression)
Joel Tscherne	Joel@tscherne.org	General
Dick Unruh	runruhjr@iw.net	Proofreading
Steve Wang	scwang@fas.harvard.edu	Statistics

A Closer Look at the OBP/SLG Ratio

Victor Wang

In a previous article, the author estimated the best coefficient for an OPS-type statistic by looking at correlation with runs scored. Here, the author updates his study by looking at runs per game, and by repeating the calculation for various run-scoring environments.

Introduction

My article published in the November 2006 edition of *By the Numbers* was an initial exploration into the correct weighting of OBP to SLG when adding the two together. By correlating different coefficients of OBP to runs scored, I discovered that weighting OBP 80% more than SLG produced the highest correlation. While the initial look into this issue did produce a result that has been confirmed by others in the sabermetric world (Hardball Times weights OBP 80% more than SLG as well in their statistic GPA), looking deeper into the data has produced some interesting results.

Corrections

It has been suggested to me that rate stats like OBP and SLG should be compared to rate stats like runs per game or runs per plate appearance to achieve the best correlations. In my initial article, I correlated the adjusted OPS to total runs. This time I correlated the adjusted OPS to runs per game, and while the conclusions are the same, the correlations I found were much higher.

Here are the old and new results. (In all tables, the highest correlation appears in bold.)

OBP Coefficient	Correlation to Runs
1.0	0.838591
1.5	0.839408
1.8	**0.840807**
1.9	0.840670
2.0	0.840491

OBP Coefficient	Correlation to Runs/Game
1.0	0.958819
1.5	0.962090
1.6	0.962311
1.7	0.962432
1.8	**0.962667**
1.9	0.962427
2.0	0.962323

As the data above shows, 1.8 is still the best coefficient for OBP when adjusting OPS. Another way to interpret this data was suggested to me by economics professor and sabermetrician Cyril Morong. Morong suggested to me that I run a multivariable regression to compare the weightings of OBP and SLG. By running a multivariable regression, I got the equation:

```
y= -4.91 + 17.12*OBP + 9.55*SLG
```

When you compare the weightings of OBP and SLG, OBP is weighted 79% more (17.12 / 9.55 = 1.79), nearly identical to the 80% weighting that was confirmed by the correlation.

Going Deeper

While using the years 1959-2006 in a correlation between adjusted OPS and R/G gives a great sample size, it also assumes that the run scoring environments for each year are similar. However, with the advancement of the HR as the primary weapon for modern baseball teams and teams scoring more runs per game, this is a false assumption. By breaking the data down into different eras, perhaps we can find different and improved weightings for OBP. I ran an initial test using the year 1985 as the break off point to see if a higher run scoring environment affects the weighting of OBP. The environments were broken into the years 1959-1984 and 1985-2006.

1959-1984

OBP Coefficient	Correlation to R/G
1.0	0.951114
1.5	0.954072
1.6	**0.954106**
1.7	0.954015
1.8	0.953819
1.9	0.953534
2.0	0.953173

1985-2006

OBP Coefficient	Correlation to R/G
1.0	0.947854
1.5	0.952490
1.6	0.952897
1.7	0.953182
1.8	0.953363
1.9	0.953453
2.0	**0.953465**
2.1	0.953409

This is very interesting information. From the years 1959-1984, the best coefficient for OBP is 1.6. However, from the years 1985-2006 the best coefficient for OBP is 2.0. By running a multivariable regression we get very similar results:

1959-1984 regression:

```
y= -4.75 + 15.96*OBP + 10.13*SLG (coefficient = 15.96/10.13 = 1.58)
```

1985-2006 regression:

```
y= -5.21 + 18.3*OBP + 9.31*SLG (coefficient = 18.3/9.31 = 1.97)
```

By breaking the years down into two eras, we find that the run scoring environment does impact the weighting of OBP. As 1985-2006 provides a higher run scoring environment, it also shows that OBP is more valuable in this kind of environment.

Choosing a Correct Break-Off Point

While these results are very encouraging, 1985 is sort of a random year in which to separate the eras. In fact, the ML average for runs per game was 4.256 in 1984, and a similar 4.331 in 1985. To find the best break-off point, we want to find a year, if any, where the ML average in runs per game sees a large jump, in the end creating almost two different run scoring environments. Hopefully such a jump exists. Fortunately, someone has already explored this topic and found a very helpful conclusion. Eric Walker, a former employee of the Oakland A's and author of *The Sinister First Baseman*, discovered in a study that the eras 1977-1992 (his data only went back to 1977) and 1994-2001 are basically two different run scoring environments.[1] While there are various reasons for this, the evidence is very conclusive. (Walker claims that the year 1993 could fit in either era but for this study we will group it with the 1994-2006 era to achieve a greater sample size.)

[1] http://highboskage.com/juiced-ball.shtml

Here are the results:

1959-1992

OBP Coefficient	Correlation to R/G
1.0	0.946879
1.5	**0.949434**
1.6	0.949393
1.7	0.949230
1.8	0.948963
1.9	0.948609
2.0	0.948181

The regression equation is

```
y= -4.69 + 15.6*OBP + 10.26*SLG (coefficient = 15.6/10.26= 1.52)
```

1993-2006

OBP Coefficient	Correlation to R/G
1.0	0.941084
1.5	0.947677
1.6	0.948309
1.7	0.948783
1.8	0.949122
1.9	0.949345
2.0	0.949468
2.1	**0.949505**
2.2	0.949468

Regression:

```
y= -5.77 + 19.79*OBP + 9.44*SLG (coefficient = 19.79/9.44 = 2.10)
```

These results really show the impact a run scoring environment has on the weighting of OBP. The era of 1959-1992 has 1.5 as the ideal weight of OBP while the era of 1993-2006 has 2.1 as the ideal weight of SLG, a weight that wasn't even included in the initial test. Walker's conclusion of two separate run scoring eras was based on data that went to the 2001 season. However, a 2007 study conducted by Ben Rader and Kenneth Winkle[2] propose three separate run-scoring eras. They deem the years 1969-1993 an era of "low productivity," the years 1994-2000 an era of "the great offensive barrage," and the years 2001-2006 an era of the "new equilibrium." If this is in fact true, based on the past findings, OBP should have lower weight than 2.1 in the current era if run scoring in this era has in fact decreased.

And so, we repeat the calculation for these three eras:

1959-1993

OBP Coefficient	Correlation to R/G
1.0	0.947802
1.5	**0.950182**
1.6	0.950115
1.7	0.949929
1.8	0.949641
1.9	0.949267
2.0	0.948822

The corresponding regression:

```
y= -4.69 + 15.52*OBP + 10.34*SLG (coefficient = 15.52/10.34 = 1.50)
```

[2] http://www.sootoday.com/content/sports/full_story.asp?StoryNumber=23819

1994-2000

OBP Coefficient	Correlation to R/G
1.0	0.945796
1.5	0.951803
1.6	0.952418
1.7	0.952896
1.8	0.953258
1.9	0.953519
2.0	0.953692
2.1	0.953790
2.2	**0.953822**
2.3	0.953798

The corresponding regression:

```
y= -5.65 + 19.99*OBP + 9.07*SLG (coefficient = 19.99/9.07 = 2.20)
```

2001-2006

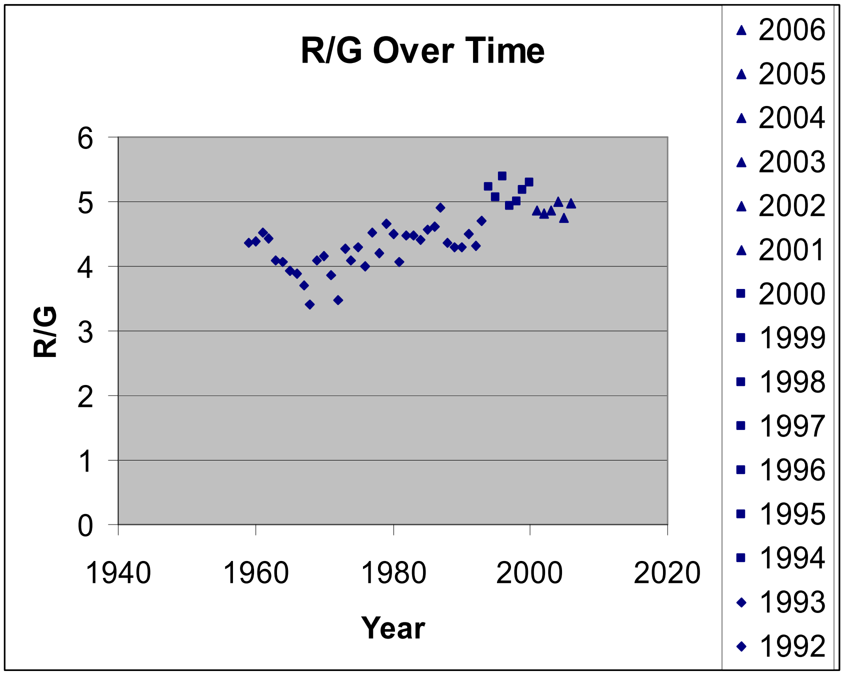
OBP Coefficient	Correlation to R/G
1.0	0.945779
1.5	0.949123
1.6	**0.949199**
1.7	0.949145
1.8	0.948981
1.9	0.948724
2.0	0.948388

Regression:

```
y= -5.68 + 17.50*OBP + 10.90*SLG (coefficient = 17.50/10.90 = 1.61)
```

Conclusion

Changing from two different eras into three different eras provides very different results. Using two eras provides a coefficient of 2.1 for OBP in today's era while breaking it down into three eras provides a coefficient of 1.6 for OBP. The reason for the drastic difference can be seen in this chart:



We previously concluded that the value of OBP increases as the offensive environment, or R/G, increases. One can identify the three different run scoring eras on the chart above (I broke the points down into different shapes to make it easier). From 1994-2000 the average R/G was 5.15 while from 2001-2006 the average R/G was 4.88. This slight decrease in R/G causes a rather large decrease in the value of OBP when compared to SLG. Interestingly enough, the value of OBP was decreasing from its peak value compared to SLG during the time OBP was becoming popularized. The most important result we can gain from this study is that the value of OBP when compared to SLG is dependent on the run scoring environment; more runs means more value for OBP. Therefore, if this "new equilibrium" run environment stays, measuring OPS by multiplying OBP by 1.8 and adding to SLG is incorrect and slightly overvalues OBP. Also, if the weighting of OBP by 1.8 becomes a sort of conventional wisdom, OBP could become overvalued when compared to SLG in today's current game if it is not already.

Victor Wang, atsbuy@yahoo.com ♦