

STATS Finds that Batters do not Get Hits When They Strikeout! Really?

J. ERIC BICKEL, PH.D

*Competitive Edge Decision Systems
6 Valleybrook Place
The Woodlands, TX 77382*

Okay, I've had it! I paid \$9.95 for my yearly subscription to STATS Inc's Fantasy Advantage and what did I learn? That Sammy Sosa is hitting .295 overall, .514 on 0-0 and only .180 with two strikes. So what? What is that supposed to tell me? That Sosa should swing at more 0-0 pitches? That Sosa can't hit with two strikes? I don't think so.

What I am about to say might shock you, but a low batting average (AVG) with two strikes does not imply that Sosa has trouble with two strikes. In fact, almost all hitters have a very low AVG with two strikes and it means almost nothing. In addition, a very high AVG on 0-0 does not imply batters should swing at more 0-0 pitches. How can this be?

AVG equals hits per at-bat (H/AB). How many ways can you have an at-bat with fewer than two strikes? Four: hit, error, fielder's choice, or batted out. The bottom line is that the ball has to be put in play. How many ways can you have an at-bat with two strikes? **Five**: hit, error, fielder's choice, batted out, *and* strikeout. The ball is either put in play or the batter strikes out.

So let's see. With less than two strikes AVG equals hits per ball put in play (BIP). Or, $AVG = H/BIP$. With two strikes AVG equals hits per BIP and strikeout (SO). Or, $AVG = H/(BIP + SO)$. Since no batter gets a hit when they strikeout, AVG with two strikes will always be lower than AVG with less than two strikes—generally much lower. Likewise, AVG on 0-0 will always be higher (generally much higher) than a batter's overall average because it eliminates the possibility of a strikeout. If you know a batter had an AB on 0-0 it could not have been a strikeout.

Maybe a simple example will help to make this clear. Let's say a batter is thrown 1000 pitches. These 1000 pitches result in the following outcomes: 308 called or swinging strikes, 35 foul balls, 405 balls, 79 hits, 8 errors, and 165 batted outs. If all these pitches were thrown with less than two strikes the AVG would be $H/(H + E + Batted\ Outs) = 79/(79 + 8 + 165) = .313$. Not bad! If these pitches were thrown with two strikes the AVG would be $H/(H + E + Batted\ Outs + Strikeouts) = 79/(79 + 8 + 165 + 308) = .141$. Wow! So this guy performs poorly with two strikes. No! In both cases he got 79 hits out of 1000 pitches or, more precisely, 79 hits out of 595 "strikes" (.132; counting balls put in play as strikes).

The moral of the story is that batters could be performing just as well (poorly) with two strikes (less than two strikes), even though their two-strike (less than two strike) AVG is low (high). Therefore, it is a fallacy to conclude that batters perform poorly (well) with two strikes (less than two strikes) simply because their batting average is low (high).

Then what does AVG by count tell you? AVG with less than two strikes tells you the fraction of balls put in play that went for hits. I call this the batter's "in play average" or IPAVG. AVG with two strikes tells you the fraction of hits that resulted from balls put in play or strikeouts. This is the standard definition of AVG. The primary reason that AVG is low with two strikes is because no batter gets a hit when they strikeout. That is really profound and certainly worth \$9.95 per year.

By the way, the same argument applies to SLG by count. Sosa is only slugging .415 with two strikes because he does not get many bases when he strikes out.

Why does STATS continue to publish these misleading stats? I'm not sure. I have to think they don't understand what AVG and SLG by count really mean. I sure wish they would stop. I am tired of Tim McCarver talking about it during every World Series like it is some profound insight. Maybe he could replace it with another statistic, like AVG by pitch type. That must be interesting.

Not so fast! Breaking down batting average by pitch type is also misleading because fewer fastballs and more off-speed pitches are thrown with two strikes. Therefore, batters will have a higher batting average off fastballs than off-speed pitches—even if they are just as likely to get a hit off of either pitch type. Based on data I have published (please see Bickel and Stotz, "Batting Average by Count and Pitch Type: Fact and Fallacy," in the 2002 *Baseball Research Journal*), AVG off fastballs and breaking balls can differ by more than 100 points—even though true batting performance is similar off either pitch type.

In summary, batting average and slugging percentage by count are highly misleading, because they imply that batters perform poorly with two strikes or incredibly well with less than two strikes. The low (high) AVG and SLG numbers with two strikes (less than two strikes) are simply defects of these statistics. This problem even affects AVG and SLG by pitch type. Unfortunately, these defects are not widely appreciated and many within baseball have been misled.

Eric Bickel earned his doctorate in decision analysis from Stanford University. Eric's primary interest is improving decision making in complex uncertain environments (e.g., baseball decisions from the field to the front office).

Competitive Edge Decision Systems provides electronic pitch/hit charting and data mining software to amateur and professional baseball and softball teams. For more information please visit www.chartmine.net or call (888) 329-0722.