The Ultimate Baseball Statistic

by Spencer M. Gluck, FCAS

With one amazing statistic, the value of each player's contribution can be measured. Power hitters versus singles hitters versus base stealers. Those who hit in the clutch versus those who choke when it really counts. Starting pitchers versus middle relievers versus closers. This statistic plays no favorites. Even the relative value of a pitcher and a hitter can be compared!

How do we do it? We take advantage of the fact that the game stops and starts. When it stops we call that a *game situation*. When the game situation changes we call that an *event*. The list of possible events is: hits, walks, hit by pitch, outs, errors, stolen bases, caught stealing, pick-offs, bases taken on throws, wild pitches, passed balls, and balks. The value of each event is defined by the effect it has on the probability of each team winning the game.[1]

---

[1] "Winning the game" is the ultimate goal we've chosen to measure against. Since we didn't pick "winning the pennant," our statistic won't reflect the fact that some games are more important than other games. That wouldn't be fair to the players whose teams are not in contention.

The value, which can be positive or negative, is charged to each team — usually to the pitcher for one team and to the batter or base runner for the other.[2] Errors are charged to neither the pitcher nor the hitter, and there are a few other exceptions we'll discuss later.

Each player winds up with a bunch of positive and negative charges — like a personal won-lost record. In fact, we'll calibrate the charges so that when a team wins, the sum total of all individual charges for its players will be 1. For a loss, the charges will add up to -1. So they really will be an individual player's share of wins and losses.

There will typically be two ways to look at these numbers: (1) *Impact,* an absolute measure equal to the sum of all the charges, positive and negative; and (2) *Effectiveness,* a percentage measure, equal to the positive charges divided by the sum of the absolute values of the positive and negative charges. It will be easy to compare players to the average, since the league average for impact will be just about 0 (that's right — below-average players will have negative impact scores), while

[2] We don't charge anything to fielding except errors. So except for errors, the pitcher gets the credit or the blame for the defense. We live within the limitations of what's captured in the official score, so we're stuck with not much on fielding. Sorry, that's the only major shortcoming we haven't solved.

the league average for effectiveness will be just about .500, or 50%.[3] The units of impact can be described as net wins or net losses. The player with the most impact in the league is probably the MVP.[4]

Now all we have to do is measure the probability of each team winning the game for every possible game situation, and we'll have the value of every event. And we have a way of doing that. But we'll save the technical details for the last section of the article, for those who are interested in such things.

We'll translate the probabilities into a *situation rating*. When the game is even (for example, when there's a tie score at the beginning of an inning), the situation rating is 0 for both teams. When the game is over, the situation rating is 1, or 100 "win points," for the winning team and -1, or 100 "loss points" for the losing team.

Better Hitting Stats

Batting average assumes that walks are meaningless and all hits are the same — but that's just not true. Slugging percentage assumes that each home run is exactly four times as valuable as each single — a cute idea but you know it's not quite right. On-base percentage says a walk's as good as a hit — but in fact, sometimes it is and sometimes

[3] If every single thing were charged to pitchers, batters, and base runners, then the averages would be exactly zero and 50%. But because of the various little exceptions, it won't be quite that exact.

[4] We don't measure fielding, and they usually like to give the MVP to a player on a team in contention. But otherwise, I think we have captured everything and I think the maximum impact player is the MVP. Even if it turns out to be a pitcher.

it’s not. Of course, a lead-off walk is exactly as good as a lead-off single, while a lead-off walk followed immediately by a steal is exactly as good as a lead-off double. With impact and effectiveness, that’s exactly how it comes out. No other stat gets it right.

And don’t forget about outs. All outs are certainly not the same. Other stats count them the same, but not so for impact and effectiveness. When you hit into an inning-ending double play with the bases loaded and the game on the line, you’ll be saddled with a negative charge equal to a bunch of unimportant outs.

Of course, there have been other attempts to capture “clutchness,” the most common being batting average with runners in scoring position. I think I’ve also seen batting average in the last three innings. How about batting average in the last three innings with runners in scoring position and a score difference of one run or less? All such crude, imperfect attempts will become unnecessary.

Better Pitching Stats

Pitching stats are in even greater need of repair. The imperfection and injustice of wins and losses I’ll just leave to the reader. The way earned runs are calculated and assigned is also loaded with problems. And saves are another grossly imperfect stat that give all glory to the guy who’s in at the end, while the poor middle relievers get no credit for anything.

Let’s look at those earned runs. When a pitcher leaves a game with two outs and a man on first, how can it be 100% his fault when that man scores? You know that’s not fair. In fact, when a reliever

comes in with men on base, he doesn’t get charged with earned runs no matter what happens, so there’s no existing stat that measures how well he does at all.

And how much is an error really worth? Let’s say that an error leaves the bases loaded with two outs when the inning would have otherwise been over. In one case, Pitcher A strikes out the next batter ending the inning, while in another case Pitcher B gives up the grand slam and five more runs after that. That’s 0 earned runs for Pitcher A and 0 earned runs for Pitcher B too. How can those performances be the same?

We’ve got all this solved. When pitchers change, all you need is a good measure of what the game situation is at the time of the change and you can divide up the responsibility fairly. And when an error causes the game situation to be different from what it would have been, just charge the difference to the error and the rest is the pitcher’s responsibility, for better or worse.

The Play by Play

As we’ve previously noted, the units of impact are net wins and losses, and we’ll call 1 net win 100 win points. Every hit, every out, every steal, every error, has an associated number of win points, and these can be reported during the course of the game. The average fan doesn’t need to understand precisely what the points are. The points correspond in such an obvious, intuitive way with the value of the hits and outs that any knowledgeable fan who hears the points assigned during the course of a game will realize that they work. At the end of the game you can identify the big blows, crucial errors, and rate the heroes and

goats of the game. And all you have to do check against what really happened in the game, and you'll know that it really works.

And now on to how we actually do it.

Calculating the Probabilities

The calculations will be based on league total statistics for a full season. For now, we'll focus on the calculations at the end of the season and worry about interim calculations later.

The charges we're calculating have to be the same for every player on every team, so when we talk about the probabilities for each team winning the game, we're not talking about any specific game. These are more like probabilities for a theoretical game between two totally average equally matched teams.

- ***The Within Half-Inning ("WHI") Situation.*** Within each half-inning, there are 24 possible situations, 3 x 8, with the three representing either 0, 1, or 2 outs, and the 8 representing nobody on; runner on first, second, or third; runner on first and second or first and third or second and third; and bases loaded. Within the course of a season for the entire league, each WHI situation will have occurred many times.

  For each occurrence of a WHI situation in the database, record the total number of additional runs scored in the half-inning. Tally them all up and we have 24 empirical distributions of additional runs scored, one for each WHI

situation. We'll use these empirical distributions just as they are.[5]

- ***The Between Half-Inning ("BHI") Situation.*** BHI situations occur at the beginning of each half-inning. They are defined by which half-inning it is and the score difference that exists at that time.

  We need to derive only one empirical distribution from the database for this calculation — the distribution of runs scored in a half-inning (this will be slightly different from the "no outs, nobody on" WHI distribution). Given a score difference at the beginning of a half-inning, we simply assume that the rest of the game will be made up of random, independent half-innings all following the same distribution of runs scored. For every possible BHI situation, it is then a simple matter to calculate the probability of each team's winning the game.

- ***Putting It All Together: The Game Situation.*** The *game situation* is defined by which half-inning it is, the score difference, and the WHI situation. Add the score difference to the additional runs-scored distribution associated with the WHI situation, and you have a distribution of BHI situations at the beginning of the next half-inning. Since we have win/lose

[5] There will be some anomalies in the "tails" of the empirical distributions. For example, you may see one case of 10 additional runs scored but no cases of 9 additional runs scored. We could fit some curve to smooth out the anomalies, but I'm against it. The anomalies in the tails will have no significant effect on our statistics. So we'll jut keep it pure and simple and use the data exactly as they come out.

probabilities associated with each BHI situation, the probability weighted average of these is the win/lose probability for the game situation.

As previously discussed, we then make a simple linear transformation to convert probabilities ranging from 0 to 1 for each team into situation ratings ranging from -1 to +1 for each team.

And that's that.

Application

The charge for each event is the change in the situation rating from before the event to after the event. The calculation of charges in most cases is pretty straightforward. All we have to do is to be sure when "before" and "after" the event is, and to whom the charge is charged.

- **Hits, walks, and outs.** In these cases, *before the event* means just before the hit, walk or out, not the beginning of the at bat. These could be different in case there was another Event during the at bat, for example a stolen base.

  After the event includes the movement of base runners who advance on the event, so a single in which a runner advances from first to third is different from one in which the runner advances only to second. Similarly, an out that advances runners is different from one that does not.[6] On

[6] Of course, we know that some base runners will advance on the same hit or out that another base runner would not advance on. So the batter gets a little of the benefit or detriment of how good the base runner is. Just a minor imperfection necessarily caused by the limitations of the official score.

the other hand, if the official score says a runner advances on a throw, then that advancement is considered a separate event, not part of the hit, walk, or out.

Hits, walks, and outs are charged to the pitcher and batter. Hit by pitch has the same effect as a walk.

- **Stolen bases, caught stealing, pick-offs.** The "before" and "after" definitions are straightforward. These are charged to the pitcher and the base runner.

- **Wild pitches, balks, bases taken on throws.** These are charged to the pitcher but not to any specific offensive player.

- **Passed balls.** These are not charged to the pitcher or to any specific offensive player.

- **Errors.** Errors present the most complicated scoring, since we must consider not two but three game situations: before, after, and as if. The as-if game situation is the one that would have occurred if there hadn't been an error. The as-if game situation will be set at the least valuable out — the out at first with the runners (if any) advancing, unless the official score says the runners advanced on the error.

  The change from the before-situation rating to the as-if-situation rating is charged to the pitcher and the batter. The difference between the as-if-situation rating and the after-situation rating is charged only to error — not

to the pitcher or to any specific offensive player.

- **Exceptions: sacrifices and walks.** When a player successfully sacrifices, he's doing what the manager requested, but it's possible that the statistics could disagree with the manager and a negative charge would result. Just in case this happens, we have a special rule that says the batter never gets a negative charge for a sacrifice. If the indicated charge is negative, he gets a 0.

  Similarly, a walk may favor the defensive team, but that's not the batter's fault. So no charge less than 0 for a walk or for being hit by a pitch either.

- **Split at bats.** A minor detail, but we're trying to cover all the bases. Sometimes the pitcher and/or batter may change during an at bat. We have no stats for split at bats, so we'll just follow the conventions of the official score. It works this way:[7] If the change occurs with less than two balls, then whatever happens is charged to the succeeding pitcher and/or batter. If the change occurs with two or more balls, then a walk is charged to the original pitcher and/or batter, while any other event is charged to the succeeding pitcher and/or batter.

Interim Calculations

[7] I think this is how it works. I'll do a little more research to make sure.

We noted earlier that the true probabilities driving the situation ratings and the charges are calculated at the end of a season based on the total stats for the season. At the beginning of a season, we'll just use the stats for the prior season to determine the situation ratings. As the season rolls on, we'll periodically recalculate the situation ratings, adding in the stats for the current season as we drop the stats for the corresponding part of the prior season. Each time we change the situation ratings, we have to recalculate all the individual player statistics for the whole season — aren't computers wonderful? We can do this as frequently as we like, but there's probably no need to do it every day. Any minor creep in the previous cumulative statistics when we recalculate will be pretty much undetectable.

And there we have it. Now we can tell exactly who the best players are without even watching the games. I think I'll keep watching the games just the same.

Bill James' response:

Well, I'm an empiricist. I start with real numbers that actually exist, and I try to see what conclusions I can draw from those numbers.

Asking me for my reaction to this proposal is very much like asking an auto mechanic what he thinks of a design for a new jet engine. I just don't know. It's above my pay grade. It might work; it might not. Build it, and I'll look it over and tell you if I can find any loose bolts. That's about all I can do.

There are very real problems here that you haven't dealt with, haven't even suggested how you would deal with. Are you going to pretend that a game situation is the same if the game is played in Coors' Field as it is if it is played in Comerica Park? If not, how are you going to account for the differences? Are you going to assume that the game situation is the same if Barry Bonds is at bat as it is if Craig Monroe is at bat? If not, how are you going to account for the differences? Are you going to assume that the game situation is the same if Juan Pierre is on base as it is if Todd Pratt is on base? If not, how are you going to account for the differences?

It is a very long and very difficult road from a fact to a conclusion. But it is a million times longer from a theory to a fact.

What you guys are trying to do here is almost the exact opposite of what I do — and, I would suggest, if you examine your own work as actuaries, you will discover that it very different from what you do, too. All good research, in my experience, starts with a question. Let me be more specific: all good research starts with a specific question which is capable of being answered:

"How often does that guy go from first to third on a single?"

"What is the value of a runner going from first to third on a single?"

"How often did that player move up on a wild pitch?"

"How often does the average player move up on a wild pitch?"

“How many times was that player thrown out on the bases last year?”

If you ask 50 questions like that, dig into the data and find the answers, then you can move on to a slightly bigger question: “Who is the best base runner on this team?” or “How good a base runner is this player?” At that point, people will pay attention to what you have to say, because you will know more about what you are talking about than anybody else does. Asking questions like this, I suspect, is analogous to what you really do in your everyday work as actuaries, and it is this work which causes you to be respected and valued by those who employ you.

On the other hand, asking a question like “Who is really the best player in baseball?” or “Who is the greatest player of all time,” as tempting as it is, does not ultimately contribute to discussion, because the true answer to the question depends upon a very large number of unknowns. We can only answer those kind of questions, in truth, by squashing all of the unknowns and then claiming that they were never important to begin with.

Good analysis never begins with statistics. Good analysis always begins with a specific question, and a question which is of interest to baseball people, whether they are actuaries or artists or aging scouts. Proposing a system which instantly evaluates everything that every player does is analogous to fixing insurance rates for drivers by attaching a box of sensors to the hood of every automobile and keeping track of how often every driver does something dangerous, and calculating exactly how dangerous that was. It’s not the real world. It’s not practical, and it’s not useful. Maybe, in 50 years, it will be practical or it will be useful, but it’s not now. Our *general* knowledge is limited by our *specific* knowledge. Our ability to

have an impact on the discussion cannot be larger than our ability to find a question which has an actual answer.