

WIKIPEDIA

The Free Encyclopedia

# Sabermetrics

In sports analytics, **sabermetrics** (originally **SABRmetrics**) is the empirical analysis of baseball, especially baseball statistics that measure in-game activity. Sabermetricians collect and summarize the relevant data from this in-game activity to answer specific questions. The term is derived from the acronym SABR, which stands for the Society for American Baseball Research, founded in 1971. The term "sabermetrics" was coined by Bill James, who is one of its pioneers and is often considered its most prominent advocate and public face.<sup>[1]</sup>

## Early history

Henry Chadwick, a sportswriter in New York, developed the box score in 1858. This was the first way statisticians were able to describe the sport of baseball by numerically tracking various aspects of game play.<sup>[2]</sup> The creation of the box score has given baseball statisticians a summary of the individual and team performances for a given game.<sup>[3]</sup>

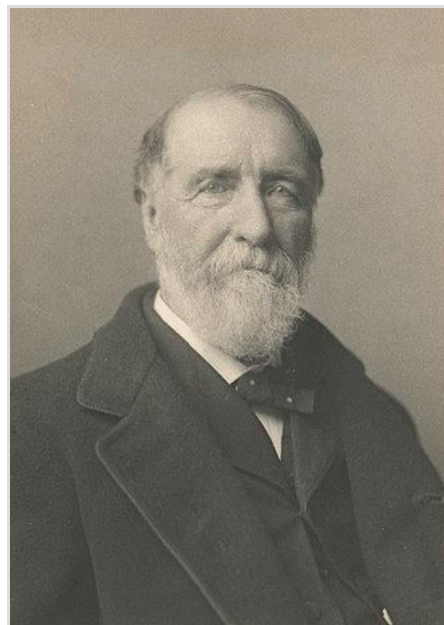
Sabermetrics research began in the middle of the 20th century with the writings of Earnshaw Cook, one of the earliest sabermetricians. Cook's 1964 book *Percentage Baseball* was one of the first of its kind.<sup>[4]</sup> At first, most organized baseball teams and professionals dismissed Cook's work as meaningless. The idea of a science of baseball statistics began to achieve legitimacy in 1977 when Bill James began releasing *Baseball Abstracts*, his annual compendium of baseball data.<sup>[5][6]</sup> However, James's ideas were slow to find widespread acceptance.<sup>[1]</sup>

Bill James believed there was a widespread misunderstanding about how the game of baseball was played, claiming the sport was not defined by its rules but actually, as summarized by engineering professor Richard J. Puerzer, "defined by the conditions under which the game is played--specifically, the ballparks but also the players, the ethics, the strategies, the equipment, and the expectations of the public."<sup>[2]</sup> Sabermetricians—sometimes considered baseball statisticians—began trying to replace the longtime favorite statistic known as the batting average.<sup>[7][8]</sup> It has been claimed that team batting average provides a relatively poor fit for team runs scored.<sup>[7]</sup> Sabermetric reasoning would say that runs win ballgames, and that a good measure of a player's worth is his ability to help his team score more runs than the opposing team.

Before Bill James popularized sabermetrics, Davey Johnson used an IBM System/360 at team owner Jerold Hoffberger's brewery to write a FORTRAN baseball computer simulation while

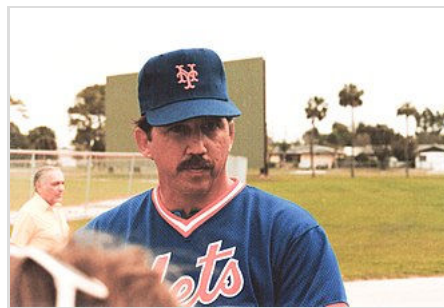


Bill James, who coined the term "sabermetrics"



Henry Chadwick

playing for the Baltimore Orioles of Major League Baseball (MLB) in the early 1970s. He used his results in an unsuccessful attempt to promote to his manager Earl Weaver the idea that he should bat second in the lineup. He wrote IBM BASIC programs to help him manage the Tidewater Tides, and after becoming manager of the New York Mets in 1984, he arranged for a team employee to write a dBASE II application to compile and store advanced metrics on team statistics.<sup>[9]</sup> Craig R. Wright was another employee in MLB, working with the Texas Rangers in the early 1980s. During his time with the Rangers, he became known as the first front office employee in MLB history to work under the title "sabermetrician".<sup>[10][11]</sup>



Davey Johnson in 1986

David Smith founded Retrosheet in 1989, with the objective of computerizing the box score of every major league baseball game ever played, in order to more accurately collect and compare the statistics of the game.

The Oakland Athletics began to use a more quantitative approach to baseball by focusing on sabermetric principles in the 1990s. This initially began with Sandy Alderson as the general manager of the team when he used the principles toward obtaining relatively undervalued players.<sup>[1]</sup> His ideas were continued when Billy Beane took over as general manager in 1997, a job he held until 2015, and hired his assistant Paul DePodesta.<sup>[8]</sup> Through the statistical analysis done by Beane and DePodesta in the 2002 season, the Oakland A's went on to win 20 games in a row. This was a historic moment for the franchise, in which the 20th game was played at the Oakland–Alameda County Coliseum.<sup>[12]</sup> His approaches to baseball soon gained national recognition when Michael Lewis published *Moneyball: The Art of Winning an Unfair Game* in 2003 to detail Beane's use of sabermetrics. In 2011, a film based on Lewis' book—also called *Moneyball*—was released and gave broad exposure to the techniques used in the Oakland Athletics' front office.



Billy Beane as a player in 1989

## Traditional measurements

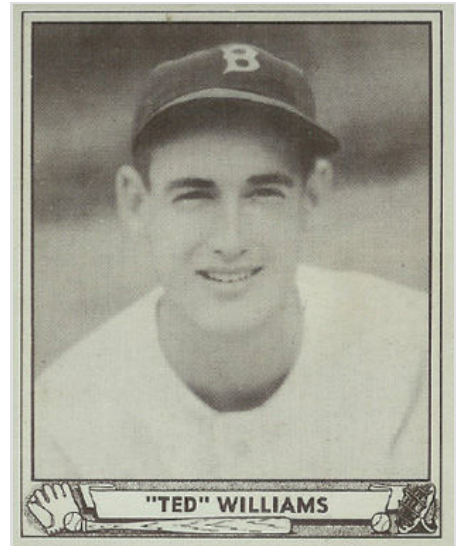
Sabermetrics was created in an attempt for baseball fans to learn about the sport through objective evidence. This is performed by evaluating players in every aspect of the game, specifically batting, pitching, and fielding. These evaluation measures are usually phrased in terms of either runs or team wins as older statistics were deemed ineffective.

### Batting measurements

The traditional measure of batting performance is considered to be hits divided by the total number of at bats. Bill James, along with other fathers of sabermetrics, found this measure to be flawed, as it ignores any other way a batter can reach base besides a hit.<sup>[13]</sup> Conversely, on-base percentage (OBP), takes base on balls ("walks") and hit-by-pitches into consideration.<sup>[a]</sup> Another issue with the traditional measure of the batting average is that it does not distinguish

between hits (i.e., singles, doubles, triples, and home runs) and gives each hit equal value.<sup>[13]</sup> A measure that differentiates among these outcomes is the slugging percentage (SLG).<sup>[a]</sup> To calculate the slugging percentage, the number of total bases of all hits is divided by the total number of times at bat.

Stephen Jay Gould proposed that the disappearance of .400 batting average (last achieved in MLB by Ted Williams in 1941) is actually a sign of general improvement in batting.<sup>[15][16]</sup> This is because, in the modern era, players are becoming more focused on hitting for power than for average.<sup>[16]</sup> Therefore, it has become more valuable to compare players using the slugging percentage and on-base percentage over the batting average.<sup>[15]</sup>



Ted Williams, the most recent MLB player to bat .400 for a season

These two improved sabermetric measures are important skills to measure in a batter and have been combined to create the modern statistic on-base plus slugging (OPS). OPS is the sum of the on-base percentage and the slugging percentage. This modern statistic has become useful in comparing players and is a powerful method of predicting runs scored from a certain player.<sup>[17]</sup>

Some of the other statistics that sabermetricians use to evaluate batting performance are weighted on-base average, secondary average, runs created, and equivalent average.

## Pitching measurements

The traditional measure of pitching performance is the earned run average (ERA). It is calculated as earned runs allowed per nine innings. Earned run average does not separate the ability of the pitcher from the abilities of the fielders that he plays with.<sup>[18]</sup> Another classic measure for pitching is a pitcher's winning percentage. Winning percentage is calculated by dividing wins by the total number of decisions (wins plus losses). Winning percentage is also heavily dependent on the pitcher's team, particularly on the number of runs it scores.



Ed Walsh, whose career 1.82 ERA is the lowest in MLB history

Sabermetricians have attempted to find different measures of pitching performance that exclude the performances of the fielders involved. One of the earliest developed, and one of the most popular in use, is walks plus hits per inning pitched (WHIP), which while not completely defense-independent, tends to indicate how many times a pitcher is likely to put a player on base (either via walk, hit-by-pitch, or base hit) and thus how effective batters are against a particular pitcher in reaching base.

A later development was the creation of defense independent pitching statistics (DIPS) system. Voros McCracken has been credited with the development of this system in 1999.<sup>[19]</sup> Through his research, McCracken was able to show that there is little to no difference between pitchers in the number of hits they allow on balls put into play—regardless of their skill level.<sup>[20]</sup> Some

examples of these statistics are defense-independent ERA, fielding independent pitching, and defense-independent component ERA. Other sabermetricians have furthered the work in DIPS, such as Tom Tango who runs the *Tango on Baseball* sabermetrics website.

*Baseball Prospectus* created another statistics called the peripheral ERA. This measure of a pitcher's performance takes hits, walks, home runs allowed, and strikeouts while adjusting for ballpark factors.<sup>[18]</sup> Each ballpark has different dimensions when it comes to the outfield wall so a pitcher should not be measured the same for each of these parks.<sup>[21]</sup>

Batting average on balls in play (BABIP) is another useful measurement for determining pitchers' performance.<sup>[20]</sup> When a pitcher has a high BABIP, they will often show improvements in the following season, while a pitcher with low BABIP will often show a decline in the following season.<sup>[20]</sup> This is based on the statistical concept of regression to the mean. Others have created various means of attempting to quantify individual pitches based on characteristics of the pitch, as opposed to runs earned or balls hit.

## Advanced methods

---

Value over replacement player (VORP) was once considered a popular sabermetric statistic. This statistic demonstrates how much a player contributes to his team in comparison to a hypothetical player that performs at the minimum level needed to hold a roster position on a major league team. This measurement was invented by Keith Woolner, a former writer for the sabermetric group/website *Baseball Prospectus*.

Wins above replacement (WAR) is another popular sabermetric statistic for evaluating a player's contributions to his team.<sup>[22]</sup> Similar to VORP, WAR compares a given player to a replacement-level player in order to determine the number of additional wins the player has provided to his team.<sup>[23]</sup> WAR values vary with hitting positions and are largely determined by a player's successful performance and amount of playing time.<sup>[23]</sup>

## Quantitative analysis in baseball

Many traditional and modern statistics, such as ERA and Wins Shared, don't give a full understanding of what is taking place on the field.<sup>[24]:189–198</sup> Simple ratios are not sufficient to understand the statistical data of baseball. Structured quantitative analysis is capable of explaining many aspects of the game, for example, to examine how often a team should attempt to steal.<sup>[25]</sup>

## Applications

---

Sabermetrics can be used for multiple purposes, but the most common are evaluating past performance and predicting future performance to determine a player's contributions to his team.<sup>[17]</sup> These may be useful when determining who should win end-of-the-season awards such as MVP and when determining the value of making a certain trade.

Most baseball players tend to play a few years in the minor leagues before they are called up to the major league. The competitive differences coupled with ballpark effects make the exact comparison of a player's statistics a problem. Sabermetricians have been able to clear this problem by adjusting the player's minor league statistics, also known as the Minor-League



Equivalency.<sup>[17]</sup> Through these adjustments, teams are able to look at a player's performance in both AA and AAA to determine if he is fit to be called up to the majors.

## Applied statistics

Sabermetrics methods are generally used for three purposes:

1. To compare key performances among certain specific players under realistic data conditions. The evaluation of past performance of a player enables an analytic overview. The comparison of this data between players can help one understand key points such as their market values. In that way, the role and the salary that should be given to that player can be defined.
2. To provide prediction of future performance of a given player or a team. When past data is available about the performance of a team or a specific player, Sabermetrics can be used to predict the average future performances for the next season. Thus, a prediction can be made with a certain probability about the number of wins and losses.
3. To provide a useful function of the player's contributions to his team. When analyzing data, one is able to understand the contributions a player makes to the success/failure of his team. Given that correlation, one can objectively sign or release players with certain characteristics.

## Machine learning model

A machine learning model can be built using data sets available at sources such as baseball-reference. This model will give probability estimates for the outcome of specific games or the performance of particular players. These estimates are increasingly accurate when applied to a large number of events over a long term. The game outcome (win/lose) is treated as having a binomial distribution.

Predictions can be made using a logistic regression model with explanatory variables including: opponents' runs scored, runs scored, shutouts, time at bat, winning rate, and pitcher whip.

## Advancements from 1985 - present

---

Bill James' two books, *The Bill James Historical Baseball Abstract* (1985) and *Win Shares* (2002) have continued to advance the field of sabermetrics.<sup>[26]</sup> His former assistant Rob Neyer, who later became a senior writer at ESPN.com and national baseball editor of SBNation, also worked on popularizing sabermetrics since the mid-1980s.<sup>[27]</sup>

Nate Silver, a former writer and managing partner of *Baseball Prospectus*, invented PECOTA in 2002–2003, introducing it to the public in the book *Baseball Prospectus* in 2003.<sup>[28]</sup> The acronym stands for *Player Empirical Comparison and Optimization Test Algorithm*,<sup>[29]</sup> and is a sabermetric system for forecasting Major League Baseball player performance. Simply put, it assumes that the careers of similar players will follow a similar trajectory. This system has been owned by *Baseball Prospectus* since 2003 and helps the website's authors invent or improve widely relied-upon sabermetric measures and techniques.<sup>[30]</sup>

Beginning in the 2007 baseball season, MLB started looking at technology to record detailed information regarding each pitch that is thrown in a game. This became known as the PITCHf/x system which is able to record the speed of the pitch, at its release point and as it crossed the

plate, as well as the location and angle of the break of certain pitches through video cameras.<sup>[13]</sup> FanGraphs is a website that favors this system as well as the analysis of play-by-play data. The website also specializes in publishing advanced baseball statistics as well as graphics that evaluate and track the performance of players and teams.

## In popular culture

- Moneyball*, the 2011 film about Billy Beane's use of sabermetrics to build the Oakland Athletics. The film is based on Michael Lewis's book of the same name.
- In the television show *Numb3rs*, the season 3 episode "Hardball" focuses on sabermetrics, and the season 1 episode "Sacrifice" also covers the subject.
- "MoneyBART", the third episode of *The Simpsons'* 22nd season, in which Lisa utilizes sabermetrics to coach Bart's Little League Baseball team.

## See also



- Analytics (ice hockey), the ice hockey equivalent
- Advanced statistics in basketball, the basketball equivalent
- Fielding Bible Award
- Kyle Boddy, founder of Driveline Baseball
- Statcast
- The Hardball Times*
- Theorycraft
- Total Baseball* by John Thorn and Pete Palmer
- Whatever Happened to the Hall of Fame?* by Bill James

## Notes

- ↑ On-base percentage and slugging percentage date to at least 1941,<sup>[14]</sup> pre-dating both Bill James (born 1949) and SABR (formed 1971).

## References

- ↑ Lewis, Michael M. (2003). *Moneyball: The Art of Winning an Unfair Game*. New York: W. W. Norton. ISBN 0-393-05765-8.
- ↑ Puerzer, Richard J. (Fall 2002). "From Scientific Baseball to Sabermetrics: Professional Baseball as a Reflection of Engineering and Management in Society". *NINE: A Journal of Baseball History and Culture*. **11**: 34–48. doi:10.1353/nin.2002.0042 (https://doi.org/10.1353%2Fnin.2002.0042). S2CID 154849268 (https://api.semanticscholar.org/CorpusID:154849268).
- ↑ "The Hall of Famers - Henry Chadwick" (https://web.archive.org/web/20080412093802/http://www.baseballhalloffame.org/hofers/detail.jsp?playerId=492558). Archived from the original (http://www.baseballhalloffame.org/hofers/detail.jsp?playerId=492558) on 2008-04-12.
- ↑ Albert, James; Jay M. Bennett (2001). *Curve Ball: Baseball, Statistics, and the Role of Chance in the Game*. Springer. pp. 170–171. ISBN 0-387-98816-5.

5. "Bill James, Beyond Baseball" (<https://www.pbs.org/thinktank/transcript1197.html>). *Think Tank with Ben Wattenberg*. PBS. June 28, 2005. Retrieved November 2, 2007.
6. Ackman, D. (May 20, 2007). "Sultan of Stats" (<http://www.opinionjournal.com/la/?id=110010232>). *The Wall Street Journal*. Retrieved November 2, 2007.
7. Jarvis, J. (2003-09-29). "A Survey of Baseball Player Performance Evaluation Measures" ([http://knology.net/~johnfjarvis/runs\\_survey.html](http://knology.net/~johnfjarvis/runs_survey.html)). Retrieved 2007-11-02.
8. Kipen, D. (June 1, 2003). "Billy Beane's brand-new ballgame" (<http://www.sfgate.com/cgi-bin/article.cgi?file=/chronicle/archive/2003/06/01/RV145326.DTL>). *San Francisco Chronicle*. Retrieved November 2, 2007.
9. Porter, Martin (1984-05-29). "The PC Goes to Bat" (<https://books.google.com/books?id=ZqO7wRZjZQQC&pg=PA210>). *PC Magazine*. p. 209. Retrieved 24 October 2013.
10. RotoJunkie – Roto 101 – Sabermetric Glossary (powered by evoArticles) (<http://www.rotokunkie.com/index.php?art/id:10>) Archived (<https://web.archive.org/web/20070910212350/http://www.rotokunkie.com/index.php?art%2Fid%3A10>) 2007-09-10 at the Wayback Machine
11. BaseballsPast.com (<http://www.baseballspast.com/radio.htm>)
12. "Franchise Timeline" (<http://oakland.athletics.mlb.com/oak/history/timeline.jsp>).
13. Albert, Jim (2010). "Sabermetrics: The Past, the Present, and the Future" (<http://www.mathaware.org/mam/2010/essays/AlbertSabermetrics.pdf>) (PDF). In Joseph A. Gallian (ed.). *Mathematics and Sports*. Vol. 43. Contributor : Mathematical Association of America. MAA. pp. 3–14. ISBN 9780883853498. JSTOR 10.4169/j.ctt6wpwsw.4 (<https://www.jstor.org/stable/10.4169/j.ctt6wpwsw.4>).
14. Powers, Jimmy (June 3, 1941). "The PowerHouse (column)" (<https://www.newspapers.com/clip/117560636/the-powerhouse-column/>). *Daily News*. New York City. p. 45. Retrieved January 30, 2023 – via newspapers.com.
15. Gould, Stephen Jay (2003). "Why No One Hits .400 Anymore" (<https://archive.org/details/triumphtragedyin00step/page/151>). *Triumph and Tragedy in Mudville: A Lifelong Passion for Baseball*. W. W. Norton & Company. pp. 151–172 (<https://archive.org/details/triumphtragedyin00step/page/151>). ISBN 0-393-05755-0.
16. Agonistas, Dan (4 August 2004). "Where have the .400 hitters gone?" (<http://danagonistes.blogspot.com/2004/08/where-have-400-hitters-gone.html>). Retrieved 30 August 2016. "... The discussion revolved around an essay that Gould wrote for *Discover* magazine in 1986 and that was reprinted both in his 1996 book *Full House* and in *Triumph and Tragedy* under the title "Why No One Hits .400 Anymore" ..."
17. Grabiner, David J. "The Sabermetric Manifesto" (<http://baseball1.com/baseball-archive/sabermetrics/sabermetric-manifesto/>). *The Baseball Archive*.
18. McCracken, Voros (January 23, 2001). "Pitching and Defense: How Much Control Do Hurlers Have?" (<http://www.baseballprospectus.com/article.php?articleid=878>). *Baseball Prospectus*.
19. Basco, Dan; Davies, Michael (Fall 2010). "The Many Flavors of DIPS: A History and an Overview". *Baseball Research Journal*. **32** (2).
20. Ball, Andrew (January 17, 2014). "How has sabermetrics changed baseball?" (<http://www.beyondtheboxscore.com/2014/1/17/5309808/how-has-sabermetrics-changed-baseball>). *Beyond the Box Score*.
21. Baumer, Benjamin; Zimbalist, Andrew (2014). *The Sabermetric Revolution: Assessing the Growth of Analytics in Baseball*. University of Pennsylvania Press.
22. Fangraphs: WAR
23. Schoenfeld, David (July 19, 2012). "What we talk about when we talk about WAR" ([http://www.espn.com/blog/sweetspot/post/\\_/id/27050/what-we-talk-about-when-we-talk-about-war](http://www.espn.com/blog/sweetspot/post/_/id/27050/what-we-talk-about-when-we-talk-about-war)). *ESPN.com*.

24. John T. Saccoman; Gabriel R. Costa; Michael R. Huber (2009). *Practicing Sabermetrics: Putting the Science of Baseball Statistics to Work*. United States of America: McFarland & Company. ISBN 978-0-7864-4177-8.
25. "The Changing Caught-Stealing Calculus | FanGraphs Baseball" (<http://www.fangraphs.com/blogs/the-changing-caught-stealing-calculus-2/>). *FanGraphs Baseball*. Retrieved 2016-12-06.
26. Neyer, Rob (November 5, 2002). "Red Sox hire James in advisory capacity" (<http://static.espn.go.com/mlb/s/2002/1105/1456563.html>). *ESPN.com*. Retrieved March 7, 2009.
27. Jaffe, C. (October 22, 2007). "Rob Neyer Interview" (<http://www.hardballtimes.com/main/article/rob-neyer-interview/>). *The Hardball Times*. Retrieved November 2, 2007.
28. Nate Silver, "Introducing PECOTA," in Gary Huckabay, Chris Kahl, Dave Pease *et al.*, Eds., *Baseball Prospectus 2003* (Dulles, VA: Brassey's Publishers, 2003): 507–514.
29. "Baseball Prospectus: Glossary" (<http://www.baseballprospectus.com/glossary/index.php?search=PECOTA>). *www.baseballprospectus.com*. Retrieved 2016-05-05.
30. "Baseball Prospectus" (<http://baseballprospectus.com/>). Retrieved 2012-03-04.

## External links

---

- [Society for American Baseball Research \(http://www.sabr.org/\)](http://www.sabr.org/) (SABR)
- 

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Sabermetrics&oldid=1211475340>"

▪