I'm very interested in learning about sabermetrics. I've found a lot of stuff on the internet through searches, but would like to cut down some time weeding through the pile. Can you please recommend some good sources that may be better suited for the basics and a beginner like myself? I plan to visit my local book store to sample the baseball section for info on sabermetrics. Thanks for any help you can provide in pointing me in the right direction.

Perfection is not attainable. But, if we chase perfection, we can catch excellence. — Vince Lombardi

---

Moneyball is a good start, as are articles by Rob Neyer at ESPN.com

www.Baseballprospectus.com is another good one. as is www.baseballthingfactory.com.

then there's always just asking questions here or at www.RedsZone.com

RIP Dimebag, Mitch, John, & Grey Cat
I would also recommend you read the articles posted at baseballthinkfactory.org and baseballprospectus.com and ask questions about what you don't understand here or within those sites...the best way to learn is to read a lot and talk to other baseball analysts...that's how I learned.

Honus Wagner Rules

xFIP?! I laugh at you!

Join Date: Nov 2004 Posts: 27419

06-23-2005, 09:11 AM

Read the Bill James Annual Baseball Abstracts. There a bit hard to find but one can learn much from them.

Strikeouts are boring! Besides that, they're fascist. Throw some ground balls - it's more democratic. -Crash Davis

SABR Matt

Hunter of Objective Truth

Join Date: May 2005 Posts: 8893

06-23-2005, 09:39 AM

I owuld say read total baseball as well if they're still publishing.

cubbieinexile

The Real McCoy

Join Date: Jan 2000 Posts: 467

06-23-2005, 10:08 AM

Go down to your Library and head for the baseball section. If you truly live in Anderson, SC then your library has a Bill James 1984 Baseball Abstract check that out plus a host of other books your library system has.

Rob Neyer would have been a great start but his archives and current stuff is no longer free.

SABR Matt

Hunter of Objective Truth

Join Date: May 2005 Posts: 8893

06-23-2005, 10:22 AM

Well that's pretty sucky...that Neyer has gone corporate...

But I understand it...I will eventually be corporate as well...sort of...I'll evenetually be selling what I have.

cubbieinexile

The Real McCoy

Join Date: Jan 2000 Posts: 467

06-23-2005, 10:36 AM
Neyer is great for beginners but outdated for anybody who has been reading his or really anybody else's stuff for awhile. Its odd most people who know Neyer and his work are vets at this and they are also the ones most likely not spend money for his articles. While on the flipside its the new guys who are most likely to enjoy his stuff yet they are not likely to buy his column since they have never had any experience with his work.

I like Neyer but his stuff stopped becoming really interesting a few years back for me.

---

**SABR Matt**
Hunter of Objective Truth

Join Date: May 2005  Posts: 8893

06-23-2005, 11:09 AM  #9

Agreed on that Cubbie.

Neyer is like a crutch for sabermetric newbies,

---

**bianchiveloce**
Registered User

Join Date: Jun 2005  Posts: 10

06-23-2005, 05:43 PM  #10

Thanks to all for the help. My interest in sabermetrics started after I read Micheal Lewis' Moneyball last week. I have a general understanding of some basic metrics like OPS, K/BB, K/9, and BB/9. I'm just now starting to read about stuff like Defense Independent metrics and PECOTA projection systems on the Baseball Prospectus site. I've also been checking out the Baseball Think Factory, The Baseball Cube, and Baseball HQ.

Thanks Cubbieinexile about the tip of checking out the library. I do indeed live in Anderson, SC. That is an excellent idea. I will pay a visit to the library before the bookstore. A great way to keep the cost down.

Also, I plan to hang around the Baseball Fever site. I found this site after doing a search for sabermetrics.

Perfection is not attainable. But, if we chase perfection, we can catch excellence. -- Vince Lombardi

---

**SABR Matt**
Hunter of Objective Truth

Join Date: May 2005  Posts: 8893

06-24-2005, 07:00 AM  #11

Go me...Go Cubbie...Go us for discussing sabermetrics often enough to get listed in online searches for that term. 😊

---

**therealnod**
Needs more cowbell

Join Date: Nov 2004  Posts: 462

06-25-2005, 01:24 PM  #12

EC,

Willie Runquist (who I was not then, nor am now familiar with) wrote an excellent book titled Baseball by the Numbers: How Statistics Are Collected, What They Mean, and How They Reveal the Game. It may require some understanding of math, but in general Runquist explains the concepts very well. I would say it is recommended, if not required, reading. I was lucky enough to have found it at my college library. For baseball books, I usually just search the library catalog for any baseball book and then go to that section and look through all the books. Don't forget reference works. My college library has two editions of Total Baseball in the reference section.

In addition to that, Baseball Prospectus has archived a series of articles for the newly interested. Then there are the HighBoskageHouse White Papers, which should give you a good idea of what, how, and why some things are done. HighBoskageHouse is Eric Walker's site, the same Eric Walker mentioned all too briefly in Moneyball as having written the "pamphlet" that Alderson gave to Beane.

I should not forget to mention Diamond Mind and Tom Tippet. Their site has a series of studies, as well. And Tangotiger, and a site called Sabernomics, which has a number of baseball studies. If you Google enough, you'll eventually find a helluva lot of sites dedicated to baseball research. Hard Ball Times, another one.
Going through all of that would see you off to a running start, I'd imagine.

Edit: Yikes! Right on the heels of SABR Matt and I forget to mention At Home Plate, which has a series of articles under the name Serious Baseball that are archived, among other things.

Last edited by therealnod; 06-25-2005, 01:30 PM.

---

**SABR Matt**  
Hunter of Objective Truth

Join Date: May 2005  Posts: 8893

06-25-2005, 01:31 PM  #13

Most of that is pretty advanced...start with the basics first. 😃

But yes, therealnod has some very good ideas for where to go when you feel comfortable with the Jamesian sabermetrics...the basic stuff.

---

**cubbieinexile**  
The Real McCoy

Join Date: Jan 2000  Posts: 467

06-26-2005, 03:00 AM  #14

Also that book is in the library in South Carolina.

---

**538280**  
Prophet of Rage

Join Date: May 2005  Posts: 11558

06-26-2005, 08:26 PM  #15

Bowling Green University has a good explanation of what sabermetrics is. You can read the whole article to find out about some basic metrics, or just read the intro to get an idea of what exactly sabermetrics is. It seems that a link won't take you to the correct place, so I'll just give you the whole article right here:

*An Introduction to Sabermetrics*

*by Jim Albert*

**What is Sabermetrics?**

Sabermetrics is the mathematical and statistical analysis of baseball records. To understand the field of sabermetrics, one first should be familiar with the game of baseball. This sport is one of the most popular games in the United States; it is often called the "national pastime." Baseball began in the eastern United States in the mid-1800's. Professional baseball started near the end of the 18th century; the National League was founded in 1876 and the American League in 1900. Currently in the United States, there are 28 professional teams in the American and National Leagues and millions of people watch games in ballparks and on television.

The game of baseball

The game of baseball is played between two teams, each consisting of nine players. The nine players are a pitcher, a catcher, first baseman, second baseman, shortstop, third baseman, left fielder, center fielder and right fielder. A game of baseball consists of nine innings. One inning is divided into two halves; in the top half of the inning, one team plays in the field and the second team comes to bat, and in the bottom half, the teams reverse roles. The team that is batting during a particular half-inning is trying to score runs. The team with the higher number of runs at the end of the nine innings is the winner of the game.

During an inning, a player on the team in the field, called a pitcher, throws a baseball toward a player of the team at-bat, called the batter. The batter will try to hit the ball using a wooden stick (called a bat) in a location out of the reach of the players in the field. By hitting the ball, the batter has the opportunity to run around four bases that lie in the field. If a player advances around all of the bases, he has scored a run. If a batter hits a ball that can be caught, or that can be thrown to first base before he runs to that base, then he is said to be out, and cannot score a run. A batter is also out if he fails to hit the baseball three times or if three good pitches (called strikes) have been thrown. The objective for the batting team during an inning is to score as many runs as possible before obtaining three outs.

The basic batting statistics

One notable aspect of the game of baseball is the wealth of numerical information that is recorded about the game. The effectiveness of batters and pitchers is typically assessed by particular numerical measures. The usual measure of hitting effectiveness for a player is the batting average which is computed by dividing the number of hits by the number of at-bats. This statistic gives the proportion of opportunities (at-bats) in which the batter succeeds (gets a hit). The batter with the highest batting average during a baseball season is called the best hitter that year. Batters are also evaluated on their ability to reach one, two,
Making adjustments

Throwing the 1919 World Series. A statistical analysis using the Player Game Percentage showed that Jackson played to his full potential during this series. Other statistics are useful in understanding pitching ability. A pitcher records a strikeout when the batter fails to hit the ball in the field and records a walk when he throws four inaccurate pitches (balls) to the batter. A pitcher who can throw the ball very fast can record a high number of strikeouts. A pitcher who is "wild" or relatively inaccurate will record a large number of walks.

Better measure of hitting ability -- runs created

One goal of sabermetrics is to find good measures of hitting and pitching performance. Bill James (1982) compares the batting records of two players, Johnny Pesky and Dick Stuart, who played in the 1960's. Pesky was a batter who hit for a high batting average but hit few home runs. Stuart, in contrast, had a modest batting average, but hit a high number of home runs. Who was the more valuable hitter? James argues that a hitter should be evaluated by his ability to create runs for his team. From an empirical study of a large collection of team hitting data, he established the following formula for predicting the number of runs scored in a season based on the number of hits, walks, at-bats, and total bases recorded in a season.

\[
\text{RUNS} = \frac{\text{HITS} + \text{WALKS}}{\text{AT-BATS} + \text{WALKS}} \times \left( \text{TOTAL BASES} \right)
\]

This formula reflects two important aspects in scoring runs in baseball. The number of hits and walks of a team reflects the team's ability to get runners on base. The number of total bases of a team reflects the team's ability to move runners that are already on base. This runs created formula can be used at an individual level to compute the number of runs that a player creates for his team. In 1942, Johnny Pesky had 620 at-bats, 205 hits, 42 walks, and 258 total bases; using the formula, he created 96 runs for his team. Dick Stuart in 1960 had 532 at-bats with 160 at-bats, 34 walks, and 309 total bases for 106 runs created. The conclusion is that Stuart in 1960 was a slightly better hitter than Pesky in 1942 since he created a few more runs for his team.

Linear weights

An alternative approach to evaluating batting performance is based on a linear weights formula. George Lindsey (1963) was the first person to assign run values to each event that could occur while a team was batting. By the use of recorded data from baseball games and probability theory, he developed the formula

\[
\text{RUNS} = \frac{\text{(HITS + WALKS)}(\text{TOTAL BASES})}{\text{AT-BATS} + \text{WALKS}}
\]

where 1B, 2B, 3B, and HR are respectively the number of singles, doubles, triples, and home runs hit in a game. One notable aspect of this formula is that it recognizes that a batter creates a run three ways. There is a direct rerun potential when a batter gets a hit and gets on base. In addition, the batter can advance runners that are already on base. Also, by not getting an out, the hitter allows a new batter a chance of getting a hit, and this produces an indirect run potential. Thorn and Palmer (1993) present a more sophisticated version of the linear weights formula which predicts the number of runs produced by an average baseball team based on all of the offensive events recorded during the game. Like James' runs created formula, the linear weights rule can be used to evaluate a player's batting performance.

Runs to wins

Although scoring runs is important in baseball, the basic objective is for a team to score more runs than its opponent. To learn about the relationship between runs scored and the number of wins, James (1982) looked at the number of runs produced, the number of runs allowed, the number of wins and the number of losses during a season for a large number of recent major league teams. James noted that the ratio of a team's wins to losses was approximately equal to the square of the ratio of runs scored to the runs allowed. Equivalently,

\[
\text{WINS} = \frac{\text{RUNS}^2}{\text{RUNS}^2 + \text{OPPOSITION RUNS}^2}
\]

where 1B, 2B, 3B, and HR are respectively the number of singles, doubles, triples, and home runs hit in a game. One notable aspect of this formula is that it recognizes that a batter creates a run three ways. There is a direct rerun potential when a batter gets a hit and gets on base. In addition, the batter can advance runners that are already on base. Also, by not getting an out, the hitter allows a new batter a chance of getting a hit, and this produces an indirect run potential. Thorn and Palmer (1993) present a more sophisticated version of the linear weights formula which predicts the number of runs produced by an average baseball team based on all of the offensive events recorded during the game. Like James' runs created formula, the linear weights rule can be used to evaluate a player's batting performance.

Better measure of pitching ability

Sabermetrics has also developed better ways of evaluating pitching ability. The standard pitching statistics, the number of wins and the earned runs per game (ERA) are flawed. The number of wins of a pitcher can just reflect the fact that he pitches for a good offensive (run scoring) team. The ERA does measure the rate of a pitcher's efficiency, but it does not tell you about the actual benefit of this pitcher over an entire season. Thorn and Palmer (1993) developed the pitching runs formula

\[
\text{PITCHING RUNS} = \text{Innings Pitched} \times \frac{\text{RUNS}}{\text{ERA}}
\]

where 1B, 2B, 3B, and HR are respectively the number of singles, doubles, triples, and home runs hit in a game. One notable aspect of this formula is that it recognizes that a batter creates a run three ways. There is a direct rerun potential when a batter gets a hit and gets on base. In addition, the batter can advance runners that are already on base. Also, by not getting an out, the hitter allows a new batter a chance of getting a hit, and this produces an indirect run potential. Thorn and Palmer (1993) present a more sophisticated version of the linear weights formula which predicts the number of runs produced by an average baseball team based on all of the offensive events recorded during the game. Like James' runs created formula, the linear weights rule can be used to evaluate a player's batting performance.

Player game percentage

Good measures of hitting, pitching, and fielding performance of baseball players have been developed. However, these statistics do not directly measure a player's contribution to a win for his team. Bennett and Flueck (1984) used data from two baseball seasons to estimate the probability the home team wins a game given the run differential (the home team runs minus visiting team runs), the half inning (top or bottom of the inning), the number of outs, and the on-base situation. Using these estimated probabilities, one can see how the probability of winning changes for each game event. One can measure a player's contribution to winning a game by summing the changes in win probabilities for each play in which the player has participated. This statistic, called the Player Game Percentage, was used by Bennett (1993) to evaluate the batting performance of Joe Jackson. This player was banished from baseball for allegedly throwing the 1919 World Series. A statistical analysis using the Player Game Percentage showed that Jackson played to his full potential during this series.

Making adjustments
Learning from selected data

Watching a baseball game raises questions that motivate interesting statistical analyses. During the broadcast of a game, a baseball announcer will typically report selected hitting data for a player. For example, it may be reported that Barry Bonds has 10 hits in his most recent 20 at-bats. What have you learned about Bonds’ batting average on the basis of this information? Clearly, Bonds’ batting average can’t be as large as 10/20 = .500 since this data was chosen to maximize the reported percentage. Casella and Berger (1994) construct the likelihood function for a player’s true batting average on the basis of this selected information and find the maximum likelihood estimate. They conclude that this selected data only provides a little insight into the “complete data” batting average that is obtained from batting records over the entire season.

Streakiness?

Another interesting question is on the existence of streakiness in hitting data. During a season it is observed that some ballplayers will experience periods of “hot” hitting where they will get a high proportion of hits. Other hitters will go through slumps or periods of hitting with very few hits. But these periods of hot and cold hitting may be just a reflection of the natural variability observed in coin tossing. Is there statistical evidence for a “hot hand” among baseball hitters where the probability of obtain a hit is dependent on recent at-bats? Albright (1993) looked at a large collection of baseball hitting data and used a number of statistics such as the number of runs to detect streakiness in hitting data. His main conclusion was that there little statistical evidence generally for a hot hand in baseball hitting.

Situational data

Currently there is great interest among fans and the media in situational baseball data. The hitting performance of batters is recorded for a number of different situations, such as day versus night games, on grass fields and artificial turf fields, against pitchers who throw right-handed and left-handed, and during home and away games. There are two basic questions in the statistical analysis of this type of data. First, are there particular situations that can explain a significant amount of variation in the hitting data? Second, are there ballplayers that perform particularly well or poorly in a given situation? Albert (1994) analyzed a large body of published situational data and used Bayesian hierarchical models to combine data from a large group of players. His basic conclusion is that there do exist some important situations. For example, batters hit on average 20 points higher when facing a pitcher of the opposite arm, and hit 8 points higher when they are playing in their home ballpark. However, there is generally little statistical evidence for individual differences in these situational effects.

Prediction

Major league baseball is currently divided into six divisions and one goal of any team is to finish first in its division. Suppose that part of the season has been completed. Using the teams’ records from this partial season, is it possible to predict accurately the winners of the divisions? Barry and Hartigan (1993) use a choice model for the probability that a team wins an individual game. This model allows for different strengths between the teams, different home advantages, and team strengths that can change randomly with time. The authors use this model to simulate the results of future baseball games and estimate the probabilities that each team will win its respective divisions.

Currently, major league baseball games are recorded in very fine detail. Information about every single ball pitched, fielded and hit during a game are noted, creating a large database of baseball statistics. This database is used in a number of ways. Public relations departments of teams use the data to publish special statistics about their players. The statistics are used to help determine the salaries of major league ballplayers. Specifically, statistical information is used as evidence in salary arbitration, a legal proceeding which sets salaries. A number of teams have employed full-time professional statistical analysts and some managers use statistical information in deciding on strategy during a game. Bill James and other baseball statisticians have shown that it is possible to answer a variety of questions about the game of baseball by means of statistical analyses.

References


The books listed in the bibliography would all be good purchases if you’re interested in the field of sabermetrics.

Last edited by 538280; 06-26-2005, 08:35 PM.
An introduction to sabermetrics, baseball analytics, data science, the R Language, and SQL. These and other areas of sabermetrics will be analyzed and better understood with current and historical baseball data. The course also serves as an applied introduction to the basics of data science, an emerging field of scholarship that requires skills in computation, statistics, and communicating results of analyses. Using baseball data, the basics of statistical regression, the R Language, and SQL will be covered. This course was successfully taught on the edX platform as a MOOC in 2014. Pursue your passion. Keep learning. Sabermetrics: An Introduction to Baseball Analysis. Discuss the game of baseball with a focus on sabermetrics and performance analytics. Take course. This course covers the theory and fundamentals of the emerging science of sabermetrics. We discuss the game of baseball, not through consensus or a fan’s conventional wisdom, but by searching for objective knowledge in baseball performance. These and other areas of baseball analytics are analyzed and better understood with current and historical baseball data. The course also serves as an applied introduction to the basics of data science, an emerging field of scholarship that requires skills in computation, statistics, and communicating results of analyses.