Teaching Statistical Thinking Using the Baseball Hall of Fame

Steve C. Wang

Statistics can be more salient to students when used in a real context. One professor at Swarthmore College used The Hall of Fame to teach statistical thinking.

Museum patrons view plaques of recent inductees into the National Baseball Hall of Fame and Museum in Cooperstown, N.Y. (AP Photo/Tim Roske)

Baseball is a natural context in which to learn about statistics. Our national pastime is replete with averages and percentages, counts and amounts, and totals of all kind. To the fan, these are not mere numbers, but condensed stories, instantly intelligible as the true literature of the game. Compared to more fluid sports such as football, basketball, hockey, and soccer, baseball especially lends itself to data collection: It is inherently sequential and discrete, with players taking turns at the plate and stopping at fixed places on the field. Such characteristics make baseball a superlative source of examples for teaching statistical principles.

In 1999, I designed and taught a course on the Baseball Hall of Fame, titled "Baseball's Highest Honor," during winter study session at Williams College. Winter study is a month-long term during January, with courses on subjects not part of the traditional academic canon. Students take one course during the winter study period, with classes typically meeting for 5–10 hours per week. Class sizes are kept small to encourage discussion.

The goal of my course was to use the Baseball Hall of Fame, particularly the process of voting for players to be enshrined, as a subject for teaching statistical and scientific principles. The course was not intended to teach students how to carry out statistical techniques, but rather to discuss what constitutes a good or bad statistical argument, and thus was in the spirit of a quantitative literacy course. Although the course was ostensibly about baseball, my underlying goal was to demonstrate the value of sound statistical and

Course Catalog Description

Math 013: Baseball's Highest Honor

Baseball's Highest Honor is election to the Hall of Fame. The history of the election process, however, has been fraught with controversy, confusion, and accusations of "favoritism." A sticking point is that there is no clear definition of what a Hall of Famer should be, other than a player of the caliber usually elected to the Hall of Fame. But with diverse standards applied by a myriad of committees, the Hall of Fame is a self-defining institution that has failed to define itself, to paraphrase author Bill James. What should the criteria be, and what, in fact, have the criteria been? How should one define "greatness"?

In this course, we will discuss the history of the Hall of Fame, methods of rating players, and election criteria. Two focuses of the course will be (1) to review modern statistical methods for evaluating players (sometimes referred to as "sabermetrics"), including models for run production and comparisons across eras, and (2) to contrast systematic, "scientific" methods for player evaluation with the ad hoc campaigning typical of Hall of Fame arguments seen today. We may also discuss the improvement or decline in quality of play over time, the relative importance of pitching and hitting, and other related issues.

scientific thinking. Essentially, this was to be a science course whose subject happened to be baseball.

The class enrolled 31 students, most of who were nonscience majors, and we met two hours per day for three afternoons every week. For each class, there were assigned readings, which included most of Bill James' 1995 book *Whatever Happened to the Hall of Fame?*, articles from various annual editions of the *Bill James Baseball Abstract*, and articles from mainstream media. Evaluation was on a High Pass/Pass/Not Pass basis and was based on class attendance and participation, a presentation, a group project, and occasional short-writing or research assignments. Finally, we ended the course by going on a class trip to Cooperstown to visit the Hall of Fame.

In baseball, the term "statistics" usually is used to refer to what academic statisticians would call "data," rather than formal inferential methods, such as hypothesis tests or confidence intervals. I retain this colloquial usage here. I also have updated all statistics through the end of the 2005 season.

Statistics in Baseball: a Sense of Distrust

Statistical arguments often are viewed with suspicion in popular culture in general and in baseball specifically. "Figures don't lie, but liars figure," people say, and who among us has not heard the old saw about "Lies, damned lies, and statistics"? In a 2005 article on *www.espn.com*, for instance, television personality Jimmy Kimmel expressed his resentment of statistics in an article about Steve Garvey's Hall of Fame qualifications:

> "We have computers now that 'revalue' baseball players of the past. They travel back in time with new and frequently nonsensical formulas designed to quantify greatness—or, more often, to make a case against it... These are the 'facts' pointed to most often when Garvey's Hall of Fame qualifications are discussed: random, machine-generated equations. His on-base percentage wasn't good enough. His OPS (whatever that is) doesn't compare to some of the other guys."

If we look at how people often use statistics in arguments about the Hall of Fame, it's easy to see why this distrust arises. In baseball (and in many other contexts), statistics too often is used as a rhetorical bludgeon to end arguments and prove points, rather than as a means of discovering knowledge. This provides an ideal opportunity to contrast such arguments with proper uses of statistics in scientific ("sabermetric") arguments. Bill James, in his 1981 *Baseball Abstract*, uses an elegant analogy:

> "Sportswriting 'analysis' is largely an adversarial process, with the most successful sportswriter being the one who is the most effective advocate of his position... Sabermetrics, by its nature, is unemotional, noncommittal. The sportswriter attempts to be a good lawyer, the sabermetrician, a fair judge."

Misuses of Statistics

One of the most common misuses of statistics is what I'll call "stat-picking": choosing only those numbers that support the writer's position and ignoring all others, while leaving out any semblance of context. In class, my students and I discussed this example from a letter to *The Sporting News* (quoted in James's *Whatever Happened to the Hall of Fame?*):

> "I often encounter the opinion that the Veterans Committee of the Baseball Hall of Fame should be disbanded because of a 'lack of worthy candidates.'
>
> "It doesn't take much homework to find a number of players worth considering. A prime example is Tony Mullane. A 30-game winner for five consecutive seasons, he won 285 games overall. He belongs in the Cooperstown shrine."

At first, the quoted numbers seem impressive—for nearly four decades no pitcher has won 30 games in a single season, let alone in five consecutive seasons. But in the 1880s, when Mullane's streak took place, starting pitchers routinely pitched in more games than they do now, recording more wins. During his five-year streak, Mullane never led the league in wins; the leaders in those seasons had 40, 43, 52, 41, and 46 wins. But these additional statistics—essential for understanding the value of Mullane's accomplishments—are omitted from the argument.

Here is another example of stat-picking from a letter to *Baseball Digest* (quoted by James) that we discussed in class:

> "He played in 1,552 games, had 5,603 at-bats, 882 runs, 1,818 hits, a .324 lifetime batting average, 181 home runs, and 997 RBI. I think he belongs in the Hall of Fame.
>
> "Here is a man who from 1926 through 1931 knocked in 90 or more runs four times and reached double figures in home runs every year. So you tell me. Why hasn't Babe Herman been elected to the Hall of Fame?"

Here again, statistics are provided without a context for understanding them. How do Herman's numbers compare with his peers? The author rattles off these statistics as if they are evidence for Herman's greatness, but are they? Herman led the league in any category only once, in triples in 1932. He knocked in 90 or more runs four times? The truly great hitters of his day were driving in upwards of 150 runs a year; the NL record of 191 RBI and the AL record of 184 RBI were both set during this period. Similarly, reaching double figures in home runs is hardly a Hall of Fame–caliber achievement.

Herman was a very good player, but as James points out, Herman often is the subject of fallacious arguments because his candidacy is a stretch. Another writer quoted by James argues that Herman "finished his 13-year major league career with a .971 fielding average, a dozen points ahead of Ty Cobb's mark." This is a remarkable statement. First of all, it's factually incorrect: Herman's fielding average was actually 10 points ahead of Cobb's, not "a dozen," but that's a quibble. Second, even if the statistic was correct, this would be evidence that Herman should not be in the Hall, not that he should. Fielding average has increased steadily over time as playing conditions and equipment have improved. The league average outfielder during Cobb's time fielded .961, and Cobb was exactly average in this respect. During Herman's career, the league average had increased to .981, so Herman's mark was actually 10 points below average. Third, and most importantly, no one thinks Cobb belongs in the Hall of Fame because of his fielding average, so the entire argument is specious. Clearly, the writer is trying to be an advocate for Herman: His goal is to marshal those statistics that bolster Herman's case, not to learn from statistics.

A related fallacy is what James labels the "in a group" argument, in which a player is compared to other players with similar statistics. The catch is that "similar" is defined only by minimum cutoffs, so the player is grouped with players at least as good as (and likely better than) he was. For example, when Rafael Palmeiro got his 3,000th hit in 2005, it was widely reported that he was only the fourth player with 3,000 hits and 500 home runs, the others being Hank Aaron, Willie Mays, and Eddie Murray. In a 2005 *Baltimore Sun* article, titled "For Palmeiro, 3,000 Hits, 500 HRs Add Up to a Spot in Hall of Fame," sportswriter Peter Schmuck argues that membership in this exclusive club should seal Palmeiro's Hall of Fame chances:

> "Rafael Palmeiro could get his 3,000th career hit at Seattle's Safeco Field tonight, and there actually is talk show debate over whether he should be a first-ballot Hall of Famer.
>
> "Please.
>
> "...I can't see how anyone could question whether he should be inducted at Cooperstown in his first year of eligibility.
>
> "Tonight, maybe tomorrow, he'll become only the fourth player in baseball history to amass both 500 home runs and 3,000 hits. The other guys are Hank Aaron, Willie Mays, and Eddie Murray. End of conversation."

Note that this was written before Palmeiro's late-2005 steroid revelations. It is literally true that only Aaron, Mays, Murray, and Palmeiro have this combination of accomplishments. But these arguments are often misleading because they invariably group the player in question with others who were better than he was. This occurs because the groups are defined by a lower bound on the number of hits or home runs, but not an upper bound. Thus, Aaron, who has 751 more hits and nearly 200 more home runs than Palmeiro, is "in the group," as is Mays with 263 more hits and nearly 100 more home runs. (Murray is a legitimately comparable player, and Schmuck does later acknowledge that Palmeiro is not in the same class as Mays or Aaron.)

People instinctively distrust such arguments because they know Herman was no Cobb and they know Palmeiro was no Mays. They too often conclude that all statistical analysis is unreliable and dishonest. And who can blame them? If you can "prove" Herman was better than Cobb, then you really can prove anything with statistics. But that's not the fault of statistics; it's the fault of an approach that uses statistics selectively to end an argument, rather than to inform an argument. Of course, there are better ways to use statistics, and comparing and contrasting these approaches is an effective way of illustrating how proper scientific methodology works. As baseball is one of the most visible contexts in which statistics is applied, recognizing the strengths of a statistical approach in baseball can go a long way toward demonstrating the value of statistics in a variety of fields.

An additional fallacy, albeit one not strictly statistical in nature, is the belief that arguing something is plausible establishes it as true. For instance, people argue that Roger Maris was aided by Yankee Stadium's short right field line in 1961, when his 61 home runs broke Babe Ruth's long-standing record. Right field in Yankee Stadium was indeed short by major league standards at 296 feet, so it is plausible this would have helped Maris, a left-handed pull hitter. But not everything plausible is true. In fact, Maris hit more home runs on the road (31) than at home (30) that year. While this is more of a logical, rather than statistical, fallacy, it does establish that, in good science, we always need to look at data, and that conclusion is certainly a statistically valuable one.

Principles of Sound Statistical Arguments

What, then, makes a statistical argument sound? In class, we discussed principles of good quantitative arguments and how these reflected principles of formal academic statistics. For instance, the "in a group" argument involves selecting subjects similar to the player in question. This argument can be made valid if the player is legitimately similar to those in the group—if, as James writes, he is "in the middle of the group." One way to select groups that are legitimately similar is to use upper and lower cutoffs, rather than just minimums. Rafael Palmeiro, for example, has 3,020 hits and 569 home runs; his comparison group might be defined by players with 2,700–3,300 hits and 450–650 home runs, rather than those with 3,000+ hits and 500+ home runs. This idea of selecting similar subjects by controlling for covariates often arises in academic statistics—in case-control studies and matched-pairs designs, for instance.

Another key principle of good statistical arguments in baseball is that statistics must include a context in which to judge them. How do we know whether "30 wins" or "90 RBI" are impressive numbers? We must ask, "relative to what?" How many games were other pitchers in the league winning at the time? How many runs were other batters driving in? Did either figure lead the league, or come close? The idea of comparison is a key principle in academic statistics as well: By itself, a treatment group is of limited use unless there is a control group with which to compare it. As Edward Tufte writes in *Visual Explanations*, "The deep, fundamental question in statistical analysis is 'Compared with what?'"

Many traditional statistics, such as pitchers' wins and RBI, are strongly dependent on factors other than the player's abilities, such as the quality of his teammates or the characteristics of his home park. Thus, arguments using these statistics are of limited value. In recent years, however, increasingly sophisticated metrics have been developed that better quantify or isolate a player's value to his team and adjust for extraneous effects. We discussed several of these in class, including on-base plus slugging, runs created, linear weights, support-neutral won-lost record, and zone rating. Academic statistics may not have direct analogues of these metrics, but the idea of isolating the effect of a factor by accounting for the effects of covariates is familiar.

These are just some of the principles for making valid arguments in baseball that have analogues in formal academic statistics. What is most important, however, is not any particular principle, but rather a general attitude toward using statistics. It is said that people most often use statistics the way a drunk uses a lamppost—for support, not for illumination.

Table 1—Comparison of Chuck Klein and Larry Walker's Seasons

	AB	HR	RBI	AVG	OBP	SLG
Klein 1932	650	38	137	.348	.404	.646
Walker 2001	497	38	123	.350	.449	.662
Klein 1933	606	28	120	.368	.422	.602
Walker 2002	477	26	104	.338	.421	.602

In contrast, the key step in formulating a valid statistical argument is to use statistics to not just prop up a teetering claim, but to shed light on the issue.

Class Assignments for Developing Sound Statistical Arguments

Because there is no self-contained definition of a Hall of Famer, two key steps in any sound Hall of Fame argument are to identify similar players who have and have not been enshrined and to compare a player to his contemporaries. Two exercises designed to clarify these tasks are Our Generation and the Keltner List. So the class would gain experience in developing sound statistical arguments, I assigned students to use these exercises to research players of their choice and present their findings to the class. Presentations usually were done in groups of three and lasted about 20 minutes.

"Our Generation" is an article written by James in his 1992 *Baseball Book* in which he compares great players of the last half-century with "similar" modern counterparts. For example, Tom Seaver is compared to Roger Clemens, and Ernie Banks to Cal Ripken. For their presentations, students repeated this exercise for some lesser-known Hall of Famers, such as Chuck Klein, Rube Marquard, and Tommy McCarthy. This was a useful way for students to get a better idea of the stature of players of previous eras. For example, students may have been vaguely familiar with Klein: They might have known he had some notable single-season marks, with as many as 250 hits and 44 assists. They might have been aware of his impressive seasons in the 1930s in a high-offense era. They might have heard his home park during those seasons, Philadelphia's Baker Bowl, was an extreme hitters' park. But was he a truly great player who benefited from his home park—like, say, Sammy Sosa—or was he merely a good player whose numbers were inflated by time and place, such as Dante Bichette? As it turns out, Klein was comparable to Larry Walker, a legitimately very good player who also played in a high-offense park and era. As of 2005, both Klein and Walker had played 17 seasons, and both were primarily right fielders. Walker's counting statistics are somewhat better, but his career park-adjusted OPS (on-base plus slugging) is close to Klein's (Walker was 40% over the league average; Klein was 37% over.). At least, superficially, Klein's 1932–33 seasons are a good match for Walker's 2001–02 seasons (see Table 1).

Most fans today have a much better conception of Larry Walker's place in the game than they do Chuck Klein's, so such comparisons to an active player are an effective way of conveying a player's stature.

The Keltner List is a series of questions intended to systematically assess a candidate's Hall of Fame qualifications. James created the list in 1984 after receiving a letter advocating Ken Keltner's enshrinement in Cooperstown because he had a higher batting average than Eddie Mathews, more RBI than Jackie Robinson, and more hits than Ralph Kiner—all three Hall of Famers. Logically, this is the equivalent of arguing that a Yugo is a great car because it's cheaper than a Rolls Royce, gets better gas mileage than a Ferrari, and is easier to parallel park than a Hummer. James instead proposed a list of 15 questions that would be more relevant to a player's Hall of Fame qualifications than "Does he have more hits than Ralph Kiner?" Here are some typical questions from the list:

> Was he ever regarded as the best player in baseball?
>
> How many MVP-type seasons did he have?
>
> Was he a good enough player that he could continue to play regularly after passing his prime?
>
> Did he have an impact on a number of pennant races?
>
> Is he the very best player in baseball history who is not in the Hall of Fame?
>
> Are most players who have comparable career statistics in the Hall of Fame?
>
> What impact did the player have on baseball history?

In their presentations, students used the Keltner List to assess the credentials of modern players of their choosing, such as Darryl Strawberry, Kirby Puckett, and Don Mattingly. Taken together, the goal of Our Generation and the Keltner List was to help students evaluate players' statistics in context—the former by comparing selected older players to modern ones and the latter by comparing selected modern players to older ones.

The course also required a final project, a 10-page paper reporting original research carried out in groups of two to four students. Any topic relating to baseball was allowed, even if it was not directly related to the Hall of Fame, as long as it could be addressed through substantive research, as opposed to opinion and conjecture. Some sample topics include:

> Are there racial or ethnic biases in Hall of Fame voting?
>
> How has the role of third basemen changed over the years? What ramifications does this have for HOF voting?
>
> What types of players are over-represented in the Hall of Fame? Among marginal selections, what traits do they have in common that are not shared by other equally worthy players who have not (yet) been selected?
>
> What types of teams exceed their runs-created estimates?

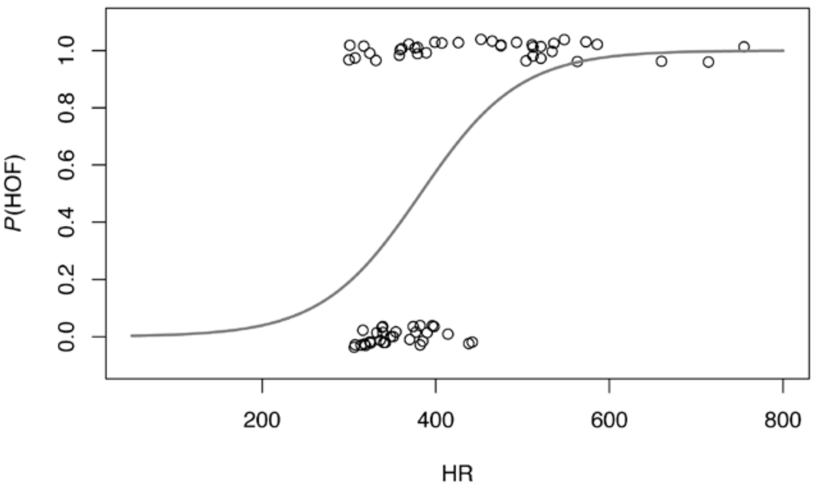


Figure 1. Best-fitting S-shaped (logistic) curve for the Baseball Hall of Fame data

> Does good pitching beat good hitting?
>
> How was the 1919 World Series viewed at the time?

The only topic I did not allow was whether player X deserves to be in the Hall of Fame. Such arguments have been rehashed at length for almost any credible candidate, and I wanted the students to do something original.

Advanced Topics: Logistic Regression

One advantage of having a class of baseball fans is that by taking advantage of this common interest, I was able to introduce substantial statistical content at a nontechnical level, including advanced topics not typically covered in introductory courses. Since all the students were familiar with the background and motivation of baseball examples, they could more easily focus on learning the statistical concepts in a context with which they were comfortable. For example, an inevitable topic in a course on the Hall of Fame is to determine the *de facto* criteria for election to the Hall of Fame. How does hitting 400 or 500 home runs affect a player's chances of being enshrined? What about having a .300 average, or 1,500 RBI? In *Whatever Happened to the Hall of Fame?*, students read about a method, the Hall of Fame Career Monitor, invented by James to answer such questions. This system involves awarding points for various achievements (e.g., eight points for winning an MVP award, five points for getting 200 hits in a season, one point for leading the league in triples, etc.). Players who reach 100 points are predicted to get into the Hall of Fame, with players above 130 almost certain to be elected. This system, however, is somewhat *ad hoc*, and the optimal parameter estimates (point values) are difficult to determine.

Here was a natural opportunity to introduce logistic regression and compare it to informal methods such as the Career Monitor. I began by asking the students to imagine making a graph of the probability that a player will be elected to the Hall of Fame (denoted by *P*(HOF)) as a function of the number of home runs he hit. I asked students to consider what such a graph would look like. For instance, would the graph be linear or curved? Students quickly determined a straight line would be nonsensical because it could predict that the probability of election exceeds 100% (for players with many home runs) or falls below 0% (for players with few home runs). Moreover, students intuitively knew that one additional home run is less important for players with many home runs (who are likely to be enshrined anyway) or those with very few home runs (who are unlikely to be enshrined, regardless), so the curve should flatten out at high and low values of home runs. On the other hand, an additional home run is most important to borderline Hall of Fame candidates, so the curve should be steepest for players on the cusp of greatness, which corresponds to about 400 HR. The students were thus able to conclude, with some guidance, that the curve should be S-shaped.

I then showed them Figure 1, which plots *P*(HOF) as a function of the number of home runs hit. I collected data from the web on all players who hit at least 300 home runs, including whether they were in the Hall of Fame. I then plotted each player's home run total (X) and HOF status (Y = 1 if in the HOF, 0 otherwise; Y-values have been jittered to improve legibility). I was not able to find systematic data listing players who hit fewer than 300 home runs, but I told the students to imagine there should be many more points representing such players in the left side of the graph, most of whom are not in the HOF. I pointed out that the curve was the "best-fitting" curve determined by a statistics program, and, indeed, it was S-shaped, with its steepest part around 400 HR.

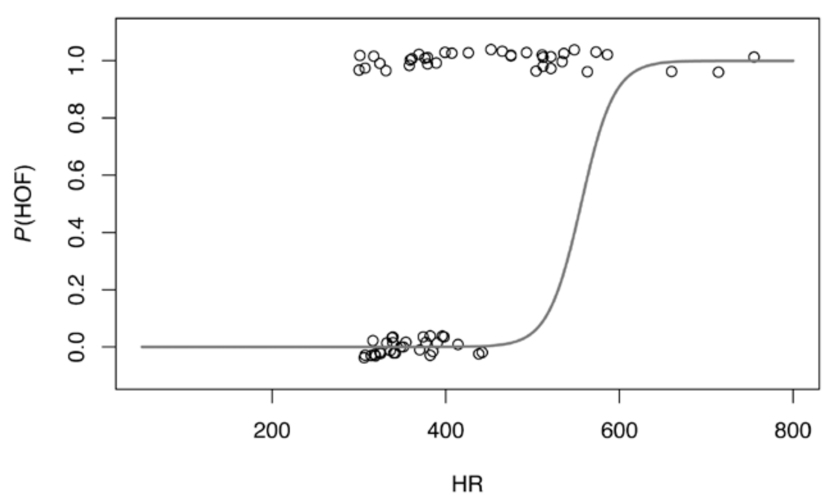


Figure 2. Poorly fitting S-shaped curve for the Baseball Hall of Fame data

I then asked the class to think about what makes this curve the "best-fitting" curve, and how one might define "best-fitting." This is not an easy question for students with little or no formal statistics background. It often is helpful to give counterexamples when introducing a new topic, so I displayed a "poorly fitting" curve as well (Figure 2). This curve clearly is shifted too far to the right; perhaps less obviously, it also is too steep.

Students clearly could see this curve did not match their intuition. For instance, it predicts that players hitting 500 HR have a miniscule chance of being enshrined when, in fact, such players are virtually guaranteed to be Hall of Famers. In other words, in light of the players who actually are or are not in the Hall of Fame, the curve in Figure 2 seems unlikely to reflect the voters' true criteria. On the other hand, the curve in Figure 1 seems much more likely in light of the actual voting. In fact, it is the most likely such curve, which is why statisticians call it the "best-fitting" curve.

I then gave the students the equation of the curve in the form

$$P(\text{HOF}) = \frac{e^{(-6.7+.0175\text{HR})}}{1+e^{(-6.7+.0175\text{HR})}}$$

Some students, particularly those with less math background, found this equation intimidating, so I rewrote the equation in a "simpler" form as

$$\frac{P(\text{HOF})}{1-P(\text{HOF})} = .00123e^{.0175HR}$$

I then asked students what the fraction on the left side represents. Sports fans typically have had enough familiarity with wagering to recognize this fraction is the odds of being elected. What this equation says, then, is that the odds of being elected is related to a player's home run total through the expression on the right side.

Students initially found the form of the right-hand expression to be cryptic. I asked them to compare the odds of election for two players who are 10 HR apart—say, 369 HR vs. 359 HR. (As it happens, these are the career HR for Ralph Kiner and Johnny Mize, respectively). The curve would predict the odds for the first player would be .00123 $e^{.0175(369)}$ = .784, whereas the odds for the second player would be .00123 $e^{.0175(359)}$ = .658. The ratio of these two odds is .784/.658 = 1.19. In other words, a player with 369 HR has 19% better odds of election than a player with 359 HR.

What about the odds of election for a player with 475 HR vs. the odds for a player with 465 HR? (These happen to be the figures for Willie Stargell and Dave Winfield.) We calculated the odds using the equation and found their ratio was again 1.19—that is, a player with 475 HR has 19% better odds of election than a player with 465 HR. By now, students were wondering if this was a coincidence, or if the ratio of the odds was always going to be 1.19 for any two players separated by 10 HR. I asked the class how we could prove the ratio will always be 1.19? We could keep trying pairs of players separated by 10 HR—perhaps 252 and 242 HR (the figures for Bobby Murcer and Dusty Baker). If we tried enough pairs and each time found a ratio of 1.19, we might be fairly confident that this relationship would always hold for any two players separated by 10 HR. But would that be proof? Is it possible, in fact, to prove this conjecture with certainty?

To answer this question, I wrote the odds ratio as $(.00123e^{.0175\,(\text{HR}+10)})/(.00123e^{.0175\,\text{HR}})$. After I cancelled terms (and briefly reviewed properties of exponents), students could see the odds ratio always would be $e^{.0175(10)}$, which equals 1.19. In other words, each 10 additional HR is associated with a 19% increase in the odds of election, regardless of the "baseline" number of HR, so we have thus proved the conjecture to be true. I asked the students whether this kind of statement sounded familiar; had they ever heard something along these lines? With some prompting, someone offered that it reminded him of medical reports: each additional pack of cigarettes you smoke increases your odds of cancer by some percent. And that, I replied, is exactly how medical researchers come up with those estimates; we've just rediscovered it in the context of baseball.

At this point, there are many options for what to cover next, depending on the mathematical level of the students. If data are available, one could apply logistic regression to a large set of predictors and compare the resulting model to James' Career Monitor. Or, one could present

Assessment

Overall, the class was a success. By the end of the course, the students were better able to critically evaluate statistical arguments. At the end, students filled out anonymous course evaluations in which they were encouraged to give honest feedback. Here are some of their comments:

Excellent. I really learned a lot. It was extremely interesting and worthwhile.

I loved it. Best winter study course I've taken in my four years. Format of class and readings was perfect.

Can I major in this? Just kidding, it was a great experience.

It was a good way to analyze baseball in a kind of scientific way and fun to learn about some of the new stats, how stats can be skewed, and to learn to distinguish good arguments from bad arguments.

It opened up a lot of doors to understanding baseball and statistics that I wasn't aware existed.

I think [the student presentations] were good, as they made us focus on a couple of specific players in order to appreciate their overall standing in the history of baseball. In doing so, we learned about the subtleties of comparing and analyzing players.

I thoroughly enjoyed the class. It gave me the means to defend my arguments with statistical evidence.

I knew enough about baseball to know I wouldn't get lost, but it taught me many new ways of looking at things and, in a very interesting way, taught me more about the intricacies of the game.

the formal statistical model underlying logistic regression, or explore concepts such as subset selection for model building. One could even discuss, at a nontechnical level, classification methods such as discriminant analysis, classification trees, and neural networks. Regardless of what one covers next, it is clearly possible to introduce important statistical concepts (conditional probability, the logistic curve, maximum likelihood estimation, odds ratios, and the nature of mathematical proof)—even for an audience with no formal statistics background—by taking advantage of the shared knowledge and intuition about baseball students bring to the course.

Additional Topics

In addition to learning about statistical and scientific reasoning, we used the Hall of Fame and baseball in general to provide a context for discussing other sociological, historical, and scientific issues.

Cities and Demographics

In discussing the effects of parks on players' statistics, we looked at photos of several old stadiums. Philadelphia's Baker Bowl, for instance, was known for inflating the batting statistics of Chuck Klein and others who played there, as noted above. The reason for Baker Bowl's generosity to hitters is clear from aerial photos: the right field line was only 281 feet deep. Indeed, the stadium looked as if a large strip of right and center fields had been lopped off. When I showed a photo of Baker Bowl, a student immediately asked why the stadium was built that way. That led to a discussion of how early 20th-century stadiums were located in cities and often had their dimensions constrained by the size and shape of the city blocks in which they were situated. Baker Bowl, for instance, had a rail yard behind the right field fence, clearly visible in aerial photos (see *www.ballparks.com*). During the 1950s and 1960s, however, stadiums often were built in the suburbs in the midst of huge parking lots, so their dimensions were unconstrained and could be made symmetrical. That led to a discussion of the American population shift from the cities to the suburbs and the ways in which baseball mirrors demographic trends.

History and Contingency

We also discussed the history of the Hall of Fame and traced the historical roots of today's controversies. For instance, the Hall of Fame originally was conceived as a museum to draw tourists to Cooperstown, which had been devastated by Prohibition and the Great Depression. Designing a selection process to honor the game's greatest players was an afterthought, and the many committees responsible for the voting have never agreed on what constitutes a Hall of Famer. (James points out that in the early years, it was not even clear what *type* of player should be elected—some advocated the enshrinement of Eddie Grant, a marginal player but a war hero who died in World War I—let alone which players should be elected.) As a result, there is no clear definition of a Hall of Famer, except a player of the caliber usually elected to the Hall of Fame. This has led to today's state of controversy, in which a multitude of candidates is debated by writers and fans using a variety of standards and agendas. Since earlier selections justify later ones, thereby compounding old errors, a different initial selection process could have led to a vastly different Hall of Fame. Such a role for contingency in history is analogous to that proposed by paleontologist Stephen Jay Gould to explain the history of life. Small changes in an ecosystem long ago, argues Gould in his book, *Wonderful Life,* have led to far-reaching evolutionary consequences millions of years later.

Scientific Revolutions

In the past few years, Michael Lewis's 2003 book, *Moneyball*, has been the subject of much controversy among baseball fans. The book describes how Billy Beane, general manager of the Oakland A's, turned to statistical analysis to build a successful team on a tight budget. *Moneyball* had not yet appeared when I was teaching my course, but if I were teaching the course today, I would certainly address it. In many ways, the process of de-emphasizing traditional scouting in favor of statistical analysis, as described in *Moneyball*, parallels the process outlined by Thomas Kuhn in his classic 1970 work on the history of science, *The Structure of Scientific Revolutions*. Kuhn argues that in scientific revolutions, such as the quantum mechanics revolution in 20th-century physics, people rarely convert from an existing paradigm (classical mechanics) to a new paradigm (quantum mechanics). Instead, the revolution is completed when a new group of scientists replaces the existing group. That process may be taking place in baseball today, as a new breed of general managers—Paul DePodesta in Los Angeles, J. P. Ricciardi in Toronto, Theo Epstein in Boston—finds statistical analysis a natural part of baseball. This so-called *Moneyball* revolution provides an excellent discussion topic: In what ways does it follow a Kuhnian scientific paradigm shift? How does it differ? And how far will the revolution go? (As I write this, DePodesta has been fired, but other young executives are making inroads.)

Conclusion

Statistical arguments about who should be in the Hall of Fame are ubiquitous. A primary goal of the course, then, was to learn to critically evaluate such arguments. What constitutes a good statistical argument or a poor one? How does a statistical or scientific argument differ from other kinds of arguments?

By the end of the course, students were able to construct sound statistical arguments and to critically think about data. In addition, they were able to identify common misuses of statistics in quantitative arguments.

Not all the students in the course were initially interested in statistics, but because students took the course by choice, all were interested in baseball—and some rabidly interested in a way not commonly encountered by teachers of introductory statistics service courses. The Baseball Hall of Fame thus provides an ideal opportunity to discuss statistical principles in a context in which many students are passionate, making statistical thinking accessible to an audience that might not otherwise take an elective course in statistics. C

Further Reading

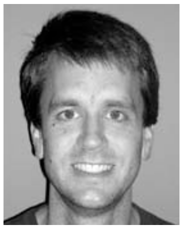
Gould, S.J. (1990). *Wonderful Life: the Burgess Shale and the Nature of History*. New York: W. W. Norton.

James, B. (2001). *The New Bill James Historical Baseball Abstract*. New York: Free Press.

James, B. (1995). *Whatever Happened to the Hall of Fame? Baseball, Cooperstown, and the Politics of Glory*. New York: Fireside.

Kuhn, T.S. (1970). *The Structure of Scientific Revolutions, 2nd Ed*. Chicago: The University of Chicago Press.

Lewis, M. (2003). *Moneyball: the Art of Winning an Unfair Game*. New York: W. W. Norton.



R. Todd Ogden earned a PhD in statistics from Texas A&M in 1994 and has worked in a variety of application areas since then. He is an associate professor in the biostatistics department of Columbia University, with a joint appointment in psychiatry. As part of this position, he works as the statistician in the Brain Imaging Division of the New York State Psychiatric Institute, a multidisciplinary group dedicated to using imaging technology to study mood disorders.

Suzanne Switzer recently earned a BA in mathematics and statistics from Smith College. Currently, she attends the Indiana University School of Library and Information Science, but when she finishes being a student, she hopes to explore the applications of statistical methods in library and information science. Until then, she's content playing both classical and Irish flute and, of course, studying. *sswitzer@smith.alumnae.net*.

Steve Wang is an assistant professor of statistics at Swarthmore College. He earned his BS from Cornell University and his MS and PhD from the University of Chicago, where he wrote his MS thesis on Markov chain models of baseball line-ups. He now works on statistical problems in paleontology. Before joining Swarthmore, he taught at Williams College and Harvard University. He became interested in statistics as a result of being a baseball fan while growing up.

Samuel S. Wu is assistant professor in the Division of Biostatistics, College of Medicine, at the University of Florida. He is also a biostatistician at the VA Rehabilitation Outcome Research Center and the VA Brain Rehabilitation Research Center in Gainesville, Florida, and a specially appointed professor in the Department of Statistics at the Tianjin University of Finance and Economics in China. *samwu@biostat.ufl.edu*.

Andrew Yang is assistant professor of Biology and Liberal Arts at the School of the Art Institute of Chicago (SAIC). His interests in biological categorization and the concept of race in genetics stems from his primary research areas of evolutionary biology and the philosophy of biology. At SAIC, he teaches various courses on biology and the interfaces between art, design, and science, including the role of visual representation in biology. He earned his PhD in biology from Duke University. **C**