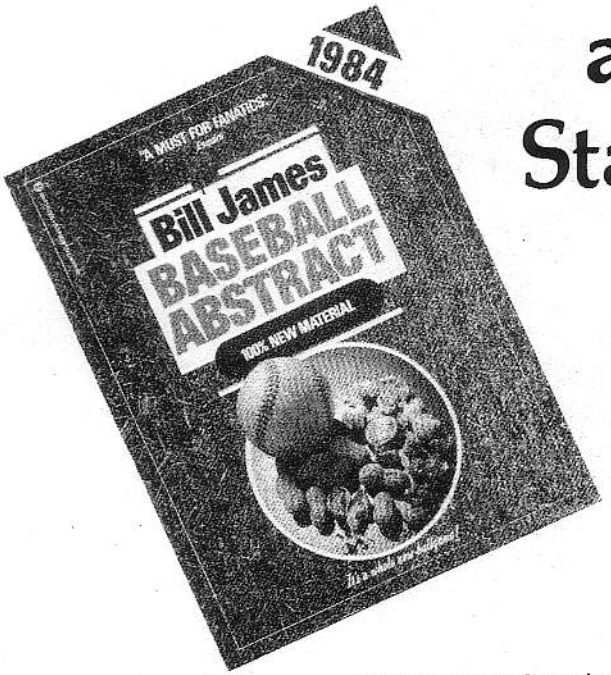
2+3 Playing with Numbers: Bill James and Using Statistics



by Robert Boone

When you write a composition, how often do you reach for a number instead of a word?

- ☐ often
- ☐ rarely
- ☐ hardly ever
- ☐ never

If you said "hardly ever" or "never," line up with the overwhelming majority of your peers. Most high school English students hunt for evidence in the familiar realm of words, not in the distant land of figures. Some will go so far as to say: "You want numbers? Go to the math class or the science lab or the computer room. But don't expect that of me in English class."

Yet writing incorporates all of life's subjects. You may, in your life after high school, be asked to use numbers in your writing as skillfully as you use words. A facility for figures will be needed in papers you write for college classes in sociology, biology, psychology, history. If you become an accountant, an insurance adjuster, a doctor, a lawyer, the manager of a dog track, or practically anything that calls for communication, you may write and read reports packed full of charts, graphs, and formulas.

If you become a professional writer, chances are you will need to blend numbers into your manuscripts far more than you probably imagine. It is true, of course, that a novelist doesn't have to worry much about using numbers, but nearly all other writers do. Of all the various writers who contribute to your daily newspaper, practically everyone must be able to →

communicate with numbers. Among this group are the sportswriters.

Sportswriters? Numbers? Add those two words together and you come up with Bill James, author of an annual statistical guide called *Baseball Abstract,* as well as hundreds of articles about his favorite subject: baseball.

Unknown a decade ago, today Bill James is the high priest of baseball statistics. Every young writer can profit from reading him. If you are irresistibly drawn to baseball, you will most certainly savor every digit of James's writing. If you are unmoved by pitchers' duels, stand-up doubles, balks, and Texas Leaguers, you will still respect James's effort. Even if you despise every single aspect of the sport, don't turn these pages in disgust—at least not until you discover how Bill James, by following a few simple rules, turns a potential weedbed of statistical jargon into a glorious garden of delights.

Use Numbers to Explain Yourself

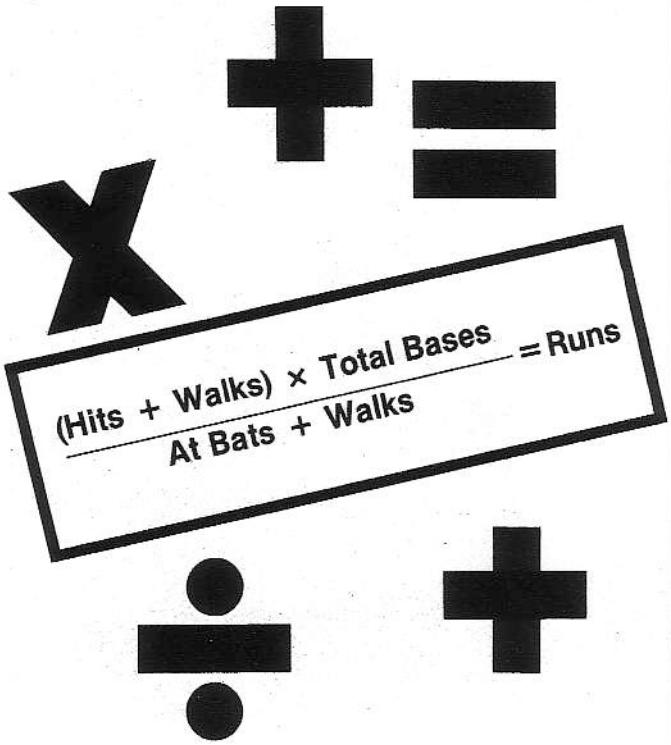
Anytime you need to use numbers to prove or illustrate a point, assume that your audience is skeptical; therefore you must always answer these two questions: Why am I using numbers at all, and how do my numerical calculations work? In the introduction of each issue of *Baseball Abstract,* James answers the first question. Here, for example, is the '84 answer:

> Statistics look at games by the hundreds, and without the details. And that is why everyone who is a baseball fan—everyone, everyone, *everyone*—reads the statistics, studies the statistics, and believes what he sees in the statistics. Without them, it is impossible to have any concept of the game, save for meaningless details floating in space.

You could leave out the numbers, he tells us, but that would kill meaningful communication.

After arguing that statistics enhance communication, he then leads us to his various numerical gadgets, explaining how each works. One of his most useful devices is to measure the true worth of a hitter with a "runs produced formula":

> Our goal now becomes to develop a formula which measures the number of runs that a player has created for his team—a formula that takes the number of hits and doubles and triples and home runs and whatever else and expresses them as runs. . . . Basic it is, simple it is. It has only four elements and requires only two simple additions, one multiplication and one division. But it works.



When he takes the reader into more bewildering mathematical regions, James still endeavors to clarify his intentions. Here, for instance, is how he uses the "runs produced formula":

> We are in a position now to answer with some accuracy the question of how many wins and how many losses result from a player's singles, doubles, triples, outs made, etc. We do this by:
>
> 1) Figuring out how many runs a player has created for his team.
>
> 2) Expressing that as a number of runs per game.

3) Adjusting for the most important illusion, the park illusion.

4) Placing this in the context of the league in which the player performs.

From this, James fashions a formula that expresses the player's offensive won/lost percentage as a decimal. Red Sox star Wade Boggs, for example, would score .778, meaning that "If every player on a team could hit the way Wade Boggs hits, and if the team had an average pitching staff and an average defense, then that team would win 77.8% of its games." Thus, James not only presents his formulas and performs the math operations; he also interprets the results. Without this explanation, much of the book would be utterly meaningless, even to the most statistically sophisticated.

Even when showing off his most advanced calculations, James still tries to make himself clear, though in certain instances, he leaves most of us readers scratching our heads:

> The Reservoir Estimation Technique assesses the "Estimated Remaining Approximate Value" or "Trade Value" of each player based on two things: his age, and his current approximate value. His age is converted into a "Y score" by the formula (24-.6 Age). The Y score and the AV are combined:

$$\frac{(AV\text{-}Y)(Y+1)\times AV}{190} \qquad \frac{AV(Y)}{13}$$

For many, the true pleasure of reading Bill James is in deciphering just such formulas, which, like profound poetry, require a large helping of concentration and patience.

Use Numbers to Answer Important Questions

"I am engaged," James says, "in a search for understanding." To help in this search, he uses numbers for the purpose of answering important (to baseball fans at least) questions such as these:

Question: How should fielding be assessed?
Answer: "According to 'range index,' which is the total number of successful plays a defensive player makes. This is much more meaningful than fielding percentage."

Question: What is the advantage of the home field?
Answer: "It is more like a proportional advantage; what it is exactly, we still don't know. Neither the straight-line assumption nor the proportional assumption can be perfectly true. . . ."

Question: What effect does a particular ballpark have on a particular batter's performance?
Answer: "It has always been my feeling that the cliquishness of the Boston Red Sox, their surliness and impatience with the press. . . was the curse of Fenway [the Red Sox park]. . . . Fenway makes ball players look like better hitters than they are. That inflates egos. Inflated egos cause resentment."

Question: What is the true value of the stolen base?
Answer: "Overrated."

Question: What effect does the American League's designated hitter rule have on a pitcher's career?
Answer: "The rate of season-to-season turnover of 200-inning pitchers is consistently higher in the American League than it is in the National, by a margin that certainly appears to be significant."

Question: What is the effect of playing at night?
Answer: "Power pitchers are more effective at night, and control pitchers more effective in the daytime."

Every once in a while, James cannot resist including answers to frivolous questions.

Frivolous question: How well do players perform on their birthdays?
Answer: [In 1983] "The birthday boys did not repeat their glorious showings of 1982."

Frivolous question: What is the composite record of the three major leaguers named "Dick Smith"?
Answer: "If there are any young men named Dick Smith out there who are entertaining notions of a career in the big leagues, some advice: either effect a name change now or plan a reevaluation of your career goals." →

Don't Attach a Number to Everything

As the two frivolous questions suggest, James has an eye for the absurd, an unlikely characteristic for one so enamored of mathematical figuring. He holds, for example, a special affection for names, such as Spike Owen (of the Seattle Mariners): "It seems a shame to waste such a good baseball nickname on somebody who is going to be out of the league in two years." James' infatuation with names sometimes results in lists such as the "all-wedding team" made up of names of real players such as McBride, Church, Parsons, Usher, Rice, and "Shot-Gun" Shuba.

"I don't cheer for numbers," James comments in the '84 *Abstract,* and he means it. Just because numbers can tell us so much about a subject like baseball, they cannot tell us everything. A true statistician like Bill James recognizes the territory impregnable to figuring.

Speak Clearly

When you read *Baseball Abstract,* you have invited Bill James into your house. Be prepared, because he will chat, rant, and amuse. He will never falsely flatter you, overwhelm you, or bore you. But he will be there, talking into your ear. Talking, talking, talking. Few writers *about any subject* achieve such a direct, clear, and honest voice.

This voice can be chatty at times, as in the introduction to the '84 *Abstract:*

> Has it been a year already? It's good to see you; lookin' good. Had a pretty good season, didn't we? A season most distinguished by the emergence of at least three players as stars of indeterminate but considerable magnitude. Looking forward to another one.
>
> I've had a good year, too; I hope the book bears the marks of it. Several things are different this year than they have been in the past. I've hired an assistant, Jim Baker; you'll be seeing his name and initials throughout the book. I bought a small home computer; it can't write and it can't think, but it has its uses. Got a dog, too. A collie.

But this cheerful voice can rise to a blustering blast as it does in his argument that Dan Quisenberry, not Lamarr Hoyt, should have won the Cy Young Award for the best pitcher in the American League:

> Shocking? Appalling, disgusting, or perhaps only ridiculous. That a whole set of voters could have reached that conclusion is bewildering, bizarre beyond description. How could it happen? What on earth could have been going on in their heads? By what reckoning did they guide their votes?

Usually, though, James' voice is simply candid and direct. Here he sums up the conclusion of the 1983 season:

> The postseason play was terrible, the worst in years. With a couple of exceptions, the games were boring, one-thrust victories almost devoid of give and take. And none of the four pennant races amounted to anything.

And he can be humorous, as in his *Sport* article assessing managers:

> I look at it this way: figuring out how to drive a car is far more complex than figuring out when to bunt and when to hit-and-run; anybody who is too stupid to manage a baseball team is never going to get to the park to begin with.

He then moves into a metaphorical voice, for humor, but also as an aid to communication. Here he is again talking about managers:

> The real value of strategy is to the fans; it's fun. So if a manager wants to be popular, the recommended fare is a bunt for breakfast, snack on stolen bases, issue two intentional walks, warm up a reliever and use a defensive substitute for lunch. Dine heartily on pitchouts and drawn-in infield, take two pinch hitters and retire early with a suicide squeeze. Make liberal use, in other words, of the full menu of strategic snack food.

Just because you are using numbers does not mean that you cannot write clearly, forcefully, and with wit. James seizes our attention by using a disarmingly direct style. He wants to be read. He not only tries to reach his audience by explaining his statistics; he also writes in unmuddied prose.

Admit Failures

Like any serious scientist, James is such a willing slave to figures that when they don't work out, rather than rationalize, he will unashamedly admit error. In the introduction to the most recent *Baseball Abstract,* he calmly states that three of his formulas from years past have been "discarded." He acknowl-

edges that his "method by which the players are rated has also been reevaluated and tuned up in light of new knowledge that has been developed."

In some instances, "new knowledge" dictated a change, while in other cases his own "shortcomings" make a change necessary: "Several people," he explains while discussing a formula, "pointed out to me, and somebody finally got it through to me, that this does not quite make sense." He even points to the limitations of his own method: "Different men, analyzing the same questions with different methods, can and do reach the same conclusions." Toward the end of his '84 *Baseball Abstract,* he analyzes the failure of his "World Series Evaluation System." He calls this chapter: "Why Did the System Have a Bad Year in 1982?"

Identify with Numbers

James has a special feeling for numbers. He describes it to Michael Lenehan in an *Atlantic Monthly* article.

> To me, all of these numbers are just like words. They all mean something. They define skills. They draw limits. The unique thing about baseball statistics is that, because of the existence of standards, they have this subtle and diverse ability to convey images.... It's quite true that they do it perhaps inaccurately, that they do it imperfectly, but nonetheless, they do it. !

Write Away

1. Consider this statement: "Numbers don't measure what really counts." Do you agree or disagree? Write a column, in newspaper editorial format, explaining your position. Use examples from your own experience, observation, and reading. Include a clear thesis statement and make your paragraphs support the main idea. (Did you notice that this question is similar in many ways to the English Achievement Test Writing Sample?)

2. Read an article that bases its conclusions on numbers. Analyze the author's use of statistics.
- What is the author trying to prove?
- What role do numbers play in this proof?
- What if numbers had been completely omitted?
- Does the author suggest any limitations to the use of numbers?
- What would happen if you took the opposite point of view? Would numbers be available to prove this position? Where might you find these numbers?
- Could the numbers have produced other conclusions?

3. Interview a writer who uses numbers regularly in his/her job. Report your findings, using either a question/answer format or the traditional expository form.
- When does this author use numbers?
- Where does he/she find these numbers?
- Has he/she developed any specific statistical skills?

4. Analyze some "pure" statistics and try to come up with original conclusions. Baseball fans, run to the *Baseball Encyclopedia.* Others, read *Information Please* or any of the other almanacs. Or, you might try the July 1984 issue of *Consumer Reports*, which has an article (loaded with numbers) about fast-food chains.

5. Consider these questions:
- How many gallons of milk have you consumed in your lifetime?
- How many miles have you walked?
- How many hours have you slept?
- How many bites have you taken?
- How many times have you yawned?
- How many hours have you studied?
- How many hours have you done nothing?

How would you go about getting answers to these questions? Compare your solutions with those of your classmates.

6. Bill James studies a world where numbers figure everywhere; yet even baseball (as he would be the first to admit) isn't totally susceptible to numerical analysis. What is your favorite activity? To what extent do statistics help answer the important questions related to this activity? Write a personal response to this question. !