The Rise and Fall of the Individual Batting Average

Peter A. Lindstrom

© The Topps Co., Inc.

Yes, it is that time of the year again – the start of the major league baseball season! Every April ignites that annual spark that lasts through World Series time in October. Although baseball is a team sport, a player's individual batting average often receives much attention from the press and the fans.

While growing up in Jamestown, NY, in the late 40's and early 50's, I was an avid baseball fan (and I still am). With the hometown Jamestown Falcons and the nearby Cleveland Indians and Pittsburgh Pirates, I often kept many batting-average statistics of the players on these teams. During the summer months, my brother Howard and I often listened to the Indians' or the Pirates' games and when a batter got a hit (or failed to get one), I would recalculate his average only if I knew the number of hits and the number of times at bat for that player.

My brother had a much easier method that did not use any of that information. If a player got a hit, he increased the players average by 2 points; otherwise, he decreased the average by 1 point. Why did Howard's method often work so well? As we develop some fundamental principles in this Pull-Out Section, we will see mathematically why his "rules of thumb" often worked so well and also when they don't work at all!

B. How a Batting Average Is Computed:

A baseball player's batting average is computed as follows:

Batting Average = Total Number of Hits/Number of Times at Bat,

where walks and sacrifices are not counted as times at bat. In the 1987 season, Wade Boggs of the Boston Red Sox had 200 hits in 551 times at bat for a 200/551 = 0.3629764, or a .363 batting average. A batting average is usually rounded off to three decimal places unless more places are necessary.

You Try It #1

Theorem 1: If x, y, a, and b are positive integers where $x/y < a/b$, then $x/y < (x + a)/(y + b) < a/b$.

Proof: If $x/y < a/b$, then we know that $(yb)(x/y) < (yb)(a/b)$, since $(yb) > 0$, or,

$$bx < ay. \quad \textbf{(A)}$$

Then,

$$(bx + ab) < (ay + ab),$$
$$b(x + a) < a(y + b),$$

so that

$$(x + a)/(y + b) < a/b. \quad \textbf{(B)}$$

Returning to **(A)**, then

$$(xy + bx) < (xy + ay),$$
$$x(y + b) < y(x + a),$$

so that

$$x/y < (x + a)/(y + b). \quad \textbf{(C)}$$

Combining **(B)** and **(C)**, we have

$$x/y < (x + a)/(y + b) < a/b.$$

a. In 1987, Tony Gwynn of the San Diego Padres had 218 hits in 589 times at bat. Determine his batting average.

b. In 1987, Robin Yount of the Milwaukee Brewers had 198 hits in 635 times at bat while Dion James of the Atlanta Braves had 154 hits in 494 times at bat. Who had the higher batting average?

Theorem 1 (see **at left**) can be used to compare batting averages.

Let's see how this theorem works with batting averages. Suppose that a major league player, early in the season, has had four hits in 12 times at bat for a 4/12 = .333 batting average. In his next game he gets three hits in four times at bat for a new batting average of (4 + 3)/(12 + 4) = 7/16 = .438. If we use $x = 4$ hits, $y = 12$ times at bat, $a = 3$ hits, and $b = 4$ times at bat, then his old batting average $= x/y < 4/12 = .333$ and $a/b = 3/4 = .750$ is his batting average for that one game. Since $x/y < a/b$, then by **Theorem 1**, $.333 = x/y < .438 = (x + a)/(y + b) < .750 = a/b$. In short, **Theorem 1** says that if a player's old batting average, x/y, is less than his average, a/b, for some period of time, then his new batting average, $(x + a)/(y + b)$, is greater than his old average and less than his average for that period of time, where x and a represent the number of hits, while y and b represent the number of times he was at bat.

You Try It #2

Theorem 2: If x, y, a, and b are positive integers where $x/y = a/b$, then $x/y = (x + a)/(y + b) = a/b$.

Theorem 3: If x, y, a, and b are positive integers where $a/b < x/y$, then $a/b < (x + a)/(y + b) < x/y$.

a. Prove each of Theorems 2 and 3 at the left.

b. Restate **Theorems 2** and **3** in terms of batting averages as was done for **Theorem 1**.

C. Changes in a Batting Average:

Theorems 1, 2, and **3** give us information about a new batting average, $(x + a)/(y + b)$, from information about an old batting average, x/y, and the batting average, a/b, for some period of time. But these theorems do not give us much information about how a player's batting average changes (increases or decreases) over that period of time.

Solution: As we have read in the fine print, our chance of winning the grand prize of $1,000,000 is about one in 32,000,000, so we can expect an average gain of (1/32,000,000)($1,000,000), which is approximately 3.1¢.

Similarly, we have a one in 320,000 chance of winning $10,000, so we expect to gain on the average (1/320,000)($10,000), which is again about 3.1¢. We have one chance in 32,000 to win $1,000, so we expect to gain (1/32,000)($1,000) on the average, which is about 3.1¢, and there is one chance in 3,200 to win $100, which gives an average expected gain of (1/3,200)($100), or about 3.1¢. Also, we have one chance in 320 to win a gift certificate worth $10, so we can expect an average gain of (1/320)($10), which again is approximately 3.1¢.

Adding these values we can expect to win 15.5¢ on the average if we enter the sweepstakes. Sounds great, but we must remember that we have to mail our entry and the stamp will cost 22¢. Therefore our net gain is 15.5¢ minus the 22¢ it costs to enter, or −6.5¢. In other words, we expect to lose 6.5¢ each time we enter the sweepstakes. This is obviously a lot different than winning a prize of $1,000,000.

You Try It #3: Suppose that you pay $1 for one of 1,000 raffle tickets, each having an equal chance to win a prize of an $800 stereo system. What is your expected gain (or loss) for this ticket?

Early information released on the proposed Iowa state lottery said that each purchaser of a $1 lottery ticket would have approximately one chance in nine of instantly winning either $2 or $5, and approximately one chance in 1,000 of winning a $100 prize in a later drawing. The Lottery Commission expected to pay out about one half of the money received in ticket sales as prize money [Daubenmier 1985].

Example #4: What can we expect to gain for each ticket we purchase in the proposed Iowa state lottery game?

Solution: In each case if we multiply the fraction which represents the probability of winning a particular prize times the value of the prize, then we get the expected value for that prize. Adding all these values gives the total expected value for the lottery game.

In this game, each entry has a 1/9 chance of instantly winning either $2 or $5, so we can represent this by the expression (1/18)($2+$5), since half of 1/9 is 1/18 and since the instant winner will win $2 half of the time and $5 half of the time. Also, there is a one in 1,000 chance of winning $100 later, so the total expected value of each entry is (1/18)($2+$5) + (1/1,000)($100). Calculated and rounded to the nearest cent, this is 49¢, so every time we pay $1 to purchase a lottery ticket we expect a return of 49¢. In the long run, we can expect to lose 51¢ every time we play!

You Try It #4: What is the expected loss for each ticket purchased if the instant winners receive either $4 or $8?

You Try It #5: A fair game is one in which the expected return is equal to the cost of entering; in other words, the expected gain or expected loss is zero. If the instant winners get either $6.20 or $10, is the lottery a fair game?

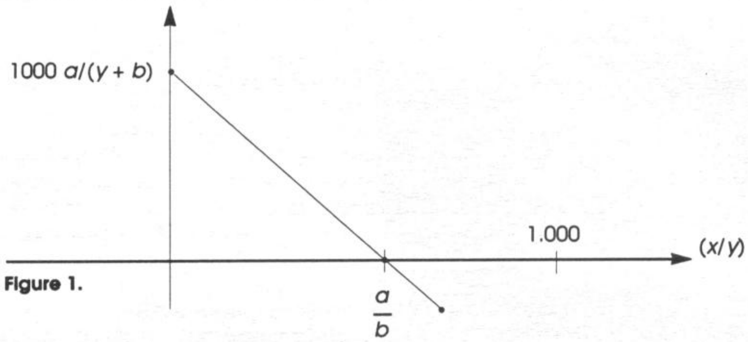


Figure 1.

From this graph, we can obtain much information:

1. If $x/y = a/b$, then $c = 0$. This same conclusion can be obtained from **Theorem 2.**
2. If $x/y < a/b$, then c is positive. This same conclusion can be drawn from **Theorem 1.**
3. If $a/b < x/y$, then c is negative. This same conclusion can be obtained from **Theorem 3.**

D. Two Special Cases of Equation (D):

Let's look at two special cases of **(D)**. In both cases, we consider what happens to c when a batter goes to bat one time.

CASE 1: Batter does not get a hit so c is negative.
Here, $a = 0$ and $b = 1$, so that **(D)** becomes

$$c = \{(-1000)/(y + 1)\}\ (x/y). \qquad \textbf{(E)}$$

From **(E)**, we can develop some specific rules for changes in a batting average depending on the number of times a player has been to bat and his old batting average.

1. When $y = 99$ times at bat and x/y is a given batting average, by using **(E)**, the change (decrease) is:

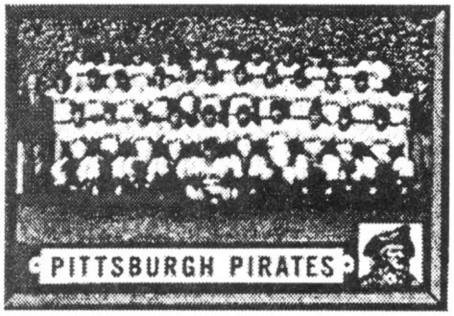
$c = ((-1000)/(99 + 1))\ (x/y),$
$c = (-10)$ (old batting average).

2. When $y = 199$ times at bat and x/y is a given batting average, by using **(E)**, the change (decrease) is:

$c = ((-1000)/(199 + 1))\ (x/y),$
$c = (-5)$ (old batting average).

You Try It #4

a. Develop similar rules for $y = 299$, 399, and 499 using **(E)**.
b. Apply an appropriate rule to estimate a player's new batting average if he presently has a batting average of .298 with 499 times at bat and he strikes out the next time he is at bat.



HOME FIELD: Forbes Field.
SEATING CAPACITY: 34,249
RIGHT FIELD FENCE: 300'
LEFT FIELD FENCE: 365'
CENTER FIELD FENCE: 457'

CASE 2: Batter gets a hit so c is positive.
Here, $a = 1$, and $b = 1$, so that **(D)** becomes

$$c = ((-1000)/(y + 1))\,(x/y) + (1000)/(y + 1),$$
$$c = (1000/(y + 1))\,(1 - x/y). \qquad \textbf{(F)}$$

From **(F)**, we can develop some specific rules for changes in a batting average depending upon the number of times a player has been to bat and his old batting average.

1. When $y = 99$ times at bat and x/y is a given batting average, using **(F)**, the change (increase) is
   $$c = ((1000)/(99 + 1))\,(1 - x/y),$$
   $$c = 10(1 - (\text{old batting average})).$$
2. When $y = 199$ times at bat and x/y is a given batting average, using **(F)**, the change (increase) is
   $$c = ((1000)/(199 + 1))\,(1 - x/y),$$
   $$c = 5(1 - (\text{old batting average})).$$

You Try It #5

a. Develop similar rules for $y = 299$, 399, and 499 using **(F)**.
b. Apply an appropriate rule to estimate a player's new batting average if he presently has a batting average of .315 with 199 times at bat and he gets a hit the next time he is at bat.

In **Table 1**, we see that the most common decrease is 1, while in **Table 2**, one of the most common increases is 2. Now I see why my brother's "rules of thumb" worked so often, but also why they worked only during the middle of the baseball season and only for certain averages!

Since the change c in both **(E)** and **(F)** depends on y, the previous number of times at bat, and the batting average x/y, **Tables 1** and **2** on the following page have been developed for various numbers of times at bat, y, and for various batting averages, x/y. **Table 1** shows the size of the decrease for Case 1 using **(E)**, and **Table 2** shows the size of the increase for Case 2 using **(F)**. All changes have been rounded to the nearest integer.

Table 1. Decreases in Batting Average

Times at Bat \ Batting Average	230	240	250	260	270	280	290	300	310	320	330	340	350	360
24	9	10	10	10	11	11	12	12	12	13	13	14	14	14
49	5	5	5	5	5	6	6	6	6	6	7	7	7	7
74	3	3	3	3	4	4	4	4	4	4	4	5	5	5
99	2	2	2	3	3	3	3	3	3	3	3	3	4	4
124	2	2	2	2	2	2	2	2	2	3	3	3	3	3
149	2	2	2	2	2	2	2	2	2	2	2	2	2	2
174	1	1	1	1	2	2	2	2	2	2	2	2	2	2
199	1	1	1	1	1	1	1	2	2	2	2	2	2	2
224	1	1	1	1	1	1	1	1	1	1	1	2	2	2
249	1	1	1	1	1	1	1	1	1	1	1	1	1	1
274	1	1	1	1	1	1	1	1	1	1	1	1	1	1
299	1	1	1	1	1	1	1	1	1	1	1	1	1	1
324	1	1	1	1	1	1	1	1	1	1	1	1	1	1
349	1	1	1	1	1	1	1	1	1	1	1	1	1	1
374	1	1	1	1	1	1	1	1	1	1	1	1	1	1
399	1	1	1	1	1	1	1	1	1	1	1	1	1	1
424	1	1	1	1	1	1	1	1	1	1	1	1	1	1
449	1	1	1	1	1	1	1	1	1	1	1	1	1	1
474	0	1	1	1	1	1	1	1	1	1	1	1	1	1
499	0	0	0	1	1	1	1	1	1	1	1	1	1	1
524	0	0	0	0	1	1	1	1	1	1	1	1	1	1
549	0	0	0	0	0	1	1	1	1	1	1	1	1	1
574	0	0	0	0	0	0	1	1	1	1	1	1	1	1
599	0	0	0	0	0	0	0	0	1	1	1	1	1	1

Table 2. Increases in Batting Average

Times at Bat \ Batting Average	230	240	250	260	270	280	290	300	310	320	330	340	350	360
24	31	30	30	30	29	29	28	28	28	27	27	26	26	26
49	15	15	15	15	15	14	14	14	14	14	13	13	13	13
74	10	10	10	10	10	10	9	9	9	9	9	9	9	9
99	8	8	8	7	7	7	7	7	7	7	7	7	7	6
124	6	6	6	6	6	6	6	6	6	5	5	5	5	5
149	5	5	5	5	5	5	5	5	5	4	4	4	4	4
174	4	4	4	4	4	4	4	4	4	4	4	4	4	4
199	4	4	4	4	4	4	4	4	3	3	3	3	3	3
224	3	3	3	3	3	3	3	3	3	3	3	3	3	3
249	3	3	3	3	3	3	3	3	3	3	3	3	3	3
274	3	3	3	3	3	3	3	3	3	2	2	2	2	2
299	3	3	3	2	2	2	2	2	2	2	2	2	2	2
324	2	2	2	2	2	2	2	2	2	2	2	2	2	2
349	2	2	2	2	2	2	2	2	2	2	2	2	2	2
374	2	2	2	2	2	2	2	2	2	2	2	2	2	2
399	2	2	2	2	2	2	2	2	2	2	2	2	2	2
424	2	2	2	2	2	2	2	2	2	2	2	2	2	2
449	2	2	2	2	2	2	2	2	2	2	1	1	1	1
474	2	2	2	2	2	2	1	1	1	1	1	1	1	1
499	2	2	2	1	1	1	1	1	1	1	1	1	1	1
524	1	1	1	1	1	1	1	1	1	1	1	1	1	1
549	1	1	1	1	1	1	1	1	1	1	1	1	1	1
574	1	1	1	1	1	1	1	1	1	1	1	1	1	1
599	1	1	1	1	1	1	1	1	1	1	1	1	1	1

E. Paul Molitor's 1987 Hitting Streak:

During the 1987 major league baseball season, Paul Molitor of the Milwaukee Brewers had a 39-game hitting streak (more baseball jargon – this means that he had at least one hit a game in 39 consecutive games). Let's see how the results of **Tables 1** and **2** can be used to estimate his batting average during part of his hitting streak. Before the start of his hitting streak, Molitor had a batting average of .323 with 50 hits in 155 times at bat. **Table 3** exhibits what he did during the first nine games of his hitting streak, along with his estimated average and his actual average. To use **Tables 1** and **2**, rounding off has to be done with regards to the number of times at bat and the average. The first estimated average is determined by using .323 as his old average.

Table 3.
First Nine Games of Molitor's 1987 Hitting Streak.

Date	Times at Bat	Hits	Decrease (Table 1)	Increase (Table 2)	Estimated Average	Actual Average
7–16–87	4	1	2, 2, 2	4	.321	51/159 = .321
7–17–87	5	3	2, 2	4, 4, 4	.329	54/164 = .329
7–18–87	4	1	2, 2, 2	4	.327	55/168 = .327
7–19–87	4	1	2, 2, 2	4	.325	56/172 = .326
7–20–87	5	3	2, 2	4, 4, 4	.333	59/177 = .333
7–22–87	3	1	2, 2	4	.333	60/180 = .333
7–23–87	5	3	2, 2	4, 4, 4	.341	63/185 = .341
7–24–87	4	1	2, 2, 2	4	.339	64/189 = .339
7–25–87	4	2	2, 2	3, 3	.341	66/193 = .342

You Try It #6

a. The following information is for the next six games of Molitor's hitting streak. Fill in the items that are missing:

Date	Times at Bat	Hits	Decrease (Table 1)	Increase (Table 2)	Estimated Average	Actual Average
7–26–87	5	3				
7–27–87	5	1				
7–28–87	3	1				
7–29–87	7	2				
7–30–87	3	2				
7–31–87	4	2				

b. State why the estimated average and the actual average can be different.

You Try It #7

a. In **Table 3**, where can you apply **Theorem 1**?
b. In **Table 3**, where can you apply **Theorem 2**?
c. In **Table 3**, where can you apply **Theorem 3**?

F. Conclusion:

Although baseball is a team sport, the individual batting average is often in the spotlight. To obtain a player's actual batting average after he gets a hit (or fails to get a hit), one must divide the number of hits by the number of times at bat. But often you do not have that information. In this Pull-Out Section, we have developed, by means of mathematics, equations and tables that enable us to estimate a player's batting average by only knowing his previous batting average and the number of times at bat.

Some Answers to the "You Try Its"

1a 218/589 = .370

b 198/635 = 0.311811 and 154/494 = 0.3117408; Yount.

2a The proofs are similar to that of **Theorem 1**.

3 $c = \{(-1000)(4)/(615 + 4)\}(205/615) + (1000)(3)/(615 + 4) = 3$ points.

4a For $y = 299$, $c = (-10/3)$ (old batting average).
For $y = 399$, $c = (-5/2)$ (old batting average).
For $y = 499$, $c = (-2)$ (old batting average).

b Using the rule for $y = 499$, $c = (-2)(.298) = -.596$, or a 1-point decrease for an estimated average of .297.

5a For $y = 299$, $c = (10/3)(1 -$ (old batting average)).
For $y = 399$, $c = (5/2)(1-$ (old batting average)).
For $y = 499$, $c = (2)(1 -$ (old batting average)).

b Using the rule for $y = 199$, $c = 5(1 - .315) = 3.425$, or a 3-point increase for an estimated average of .318.

6a

Date	Times at Bat	Hits	Decrease (Table 1)	Increase (Table 2)	Estimated Average	Actual Average
7–26–87	5	3	2, 2	3, 3, 3	.346	69/198 = .348
7–27–87	5	1	2, 2, 2, 2	3	.341	70/203 = .345
7–28–87	3	1	2, 2	3	.340	71/206 = .345
7–29–87	7	2	2, 2, 2, 2, 2	3, 3	.336	73/213 = .343
7–30–87	3	2	2	3, 3	.340	75/216 = .347
7–31–87	4	2	2,2	3, 3	.342	77/220 = .350

b Rounding off, **Tables 1** and **2** have times at bat expressed in increments of 25 and batting averages expressed in increments of 10 points; smaller increments would produce more accurate estimates.

7a On 7–17–87, 7–20–87, 7–23–87, and 7–25–87.

b On 7–22–87.

c On 7–16–87, 7–18–87, 7–19–87, and 7–24–87.