

The flaw in batting averages



Sanchit Maini and Sumit Narayanan point out the flaw in batting averages that stops Viv Richards from having a higher batting average than Michael Bevan.



So, can you tell the true batting averages of all the great and not-so-great one-day players? They're all published and publicised, quoted, and misquoted. But does the batting average as we know it truly reflect a measure of central tendency? We don't think so.

So, what's wrong?

What's wrong with the current method of calculating averages is the treatment of 'not-outs'. The current definition of batting averages takes the total numbers of runs scored by players and divides it by the number of innings played. However, the denominator excludes the number of innings in which the batsman was not out. The logic for ignoring the not-out innings is that it is not possible to estimate the number of runs a batsman would have made had the game continued. However, this really bumps up the career batting average for late-order batsmen like Michael Bevan.

All those versed with the concept of exposed-to-risk will know there's something wrong here. For example, if a batsman scores 300 runs in four innings of a series and is not out in each of them, his average will be infinite. In the fifth innings if he gets out for a duck, his average for the series would be a majestic 300!

Clearly, batsmen who do not have a substantial number of not-outs in their careers would not be affected much, but those late-order batsmen who stay around until the end of the innings would benefit unduly under the current method of calculating batting average.

An alternative measure

An alternative to the current method is to account for, using some proxy, the innings in which the batsman was not out. One proxy can be the career average number of minutes spent by a batsman in the middle. The problem with this proxy is that a batsman could have spent a lot of minutes in the middle, but on the non-striker's end.

Another and a better proxy (we assert) is the career average number of balls faced per innings. This essentially uses the principle of exposed-to-risk for calculating the denominator in the average. Each inning of a player can be viewed as contributing to the 'exposure' and the fact that a batsman ends up being not out at the end of the innings does not indicate that the exposure was zero, since the player has had the opportunity to score runs in that innings.

We propose the following formula to calculate the following exposure:



Sanchit Maini is vice-president, actuarial services at Max New York Life in India

```

IF Batsman = Out
THEN 1
ELSEIF Batsman = Not Out
AND NOB<AvgNOB
THEN NOB/AvgNOB
ELSE 1
ENDIF

```

where,
 NOB = Number of balls faced in a particular innings
 AvgNOB = Career average number of balls faced per innings

Therefore, for an innings in which the batsman is not out we can define the exposure as the actual number of balls faced in the innings over the average number of balls faced by the batsman in his career. This exposure is capped at one, so if the batsman is not out but has faced more than the average number of balls then he has played a 'full' innings for the purpose of calculating the batting average.

Does this really matter?

This depends on whether it's Tendulkar or Bevan, Sehwag, or Klusener. Top-order batsmen are less likely to have many not-outs in their career, so the revised batting average will not be very different from the normal average. However, middle-order and lower middle-order batsmen, especially those that stick around to the end, are the ones most affected. Michael Bevan, the quintessential 'finisher' in one-dayers, ended with a phenomenal average of 53.58 in 232 ODIs. However, his revised average drops to 38.73!

Similarly, Lance Klusener's average drops from 41.10 to 28.73. In our sample of batsmen, Virender Sehwag shows the maximum resistance and his average only drops from 32.44 to 31.20. We all know the reason why: he's not very likely to remain not out at the end of 50 overs!

Do the results make sense?

Table 1 shows the results for the sample batsmen we looked at sorted by the drop in their averages. The top 10 batsmen in the table whose averages have dropped the most are all middle-order batsmen. Of these, five are Aussies.

An alternative, and simpler way of calculating the revised average is to deduct from the normal average the proportion of not-outs to matches. This gives the estimate (shown as 'revised average (2)' in table 1) that has an error on average of less than half a run. The method gives an estimate of the average score that is reasonable.

Comparing the two averages with the median

Figure 1 Comparison of averages

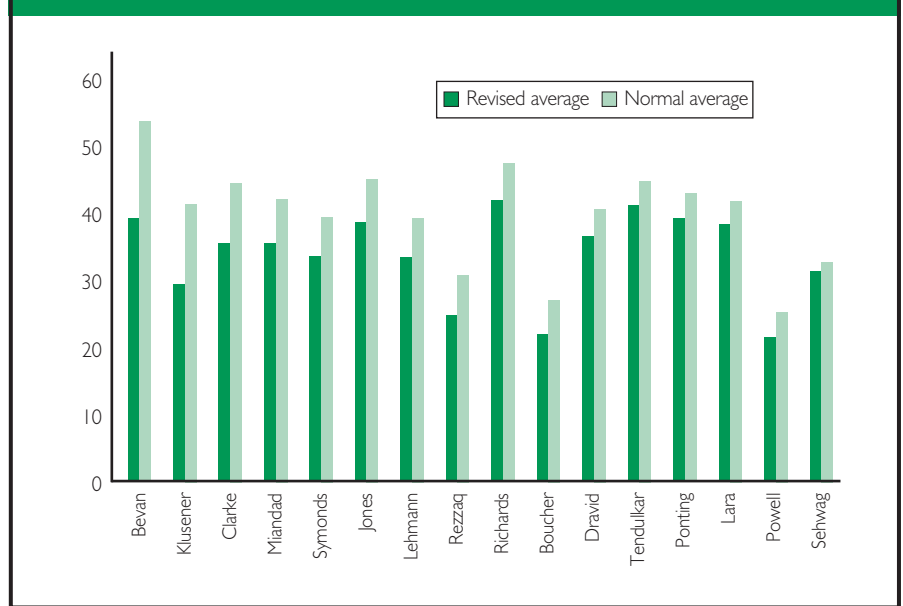


Table 1 Sample batsmen sorted by drop in averages

Player	Country	Matches	Runs	Not out	Normal average	Revised average(1)	Difference	Not out /matches	Revised average(2)	Error
Bevan (4)	Aus	232	6,912	67	53.6	38.7	14.9	28.9%	38.1	0.6
Klusener (13)	SA	171	3,576	50	41.1	28.9	12.3	29.2%	29.1	-0.2
Clarke (9)	Aus	82	2,389	19	44.2	34.7	9.6	23.2%	34.0	0.7
Miandad (8)	Pak	233	7,381	41	41.7	35.1	6.6	17.6%	34.4	0.8
Symonds (11)	Aus	146	3,697	22	39.3	32.8	6.5	15.1%	33.4	-0.6
Jones (5)	Aus	164	6,068	25	44.6	38.2	6.4	15.2%	37.8	0.4
Lehmann (10)	Aus	117	3,078	22	39.0	33.0	6.0	18.8%	31.6	1.3
Rezaq (14)	Pak	212	4,201	46	30.4	24.7	5.7	21.7%	23.8	0.9
Richards (1)	WI	187	6,721	24	47.0	41.4	5.6	12.8%	41.0	0.4
Boucher (15)	SA	217	3,175	39	26.7	21.9	4.8	18.0%	21.9	0.0
Dravid (7)	Ind	292	9,528	34	40.2	36.1	4.1	11.6%	35.5	0.6
Tendulkar (2)	Ind	362	14,146	33	44.2	40.2	3.8	9.1%	40.2	0.3
Ponting (3)	Aus	252	9,210	29	42.4	38.7	3.7	11.5%	37.6	1.2
Lara (6)	WI	270	9,661	27	41.1	37.7	3.4	10.0%	37.0	0.7
Powell (16)	WI	109	2,085	16	24.8	21.6	3.2	14.7%	21.2	0.4
Sehwag (12)	Ind	152	4,607	7	32.4	31.2	1.2	4.6%	30.9	0.3

Note: Figure in brackets is the rank according to the revised average. All ODIs played until June 2006 have been included in the analysis.

score also illustrates that the revised average is closer to the median than the normal average. However, the median is still lower than both the revised and the normal average. This indicates that the distribution of runs scored by batsmen is positively skewed – that is, there are a few high scores above the median which push the averages up.

We believe that our proposed adjustment to the current methodology for calculating batting average has some advantages and is a better reflection of the underlying central tendency. In our sample, Viv Richards has the second-highest

normal average. Under the revised method, Richards moves to the top, with Sachin Tendulkar moving to the second position and Ricky Ponting at 3. That sounds about right! □



Sumit Narayanan is a consulting actuary at Watson Wyatt in Singapore