Performance Analysis

Run Crafting Insights and Stroke Play

Batting Performance Analysis

In every format of cricket, there are couple of metrics available to measure batting performance of a cricketer. Among them, the most used and popular is Batting Average and Strike Rate.

Batting Average is defined by the ratio of the total runs scored by a batter to the total number of dismissed innings (i.e. the number of innings in which he/she got dismissed by the opponent teams).

Mathematically, we can write the Batting Average (Traditional Batting Average) as

$$\text{TBA} = \frac{\text{Total Runs Scored}}{\text{Total Number of Dismissed Innings}}$$

One of the main reasons the batting average is so widely used is its simplicity. It serves as an indicator of a bater's performance over a particular period or series and is often used to compare different players. These comparisons can be based on overall performance or specific conditions such as against a particular team or at a certain venue. Unlike total runs scored, the batting average provides a more meaningful metric for evaluating a batter's performance over both short durations—like a series or tournament—and longer spans, such as a year or an entire career.

The statistical foundation for making such comparisons rests on the assumption that the batting average, as calculated, serves as a reliable estimator of the true average of the underlying population. In other words, it assumes that the number of runs a batter scores in a randomly selected match within a given context reflects this true average.

Let us consider an example of batting average unrepresentative of the true average performance in a few matches, which is also highly overestimated measurement in this context.

Suppose a batter has played 10 innings and scored $x_1, x_2, \dots, x_{10}$ runs. Among those, he was dismissed only one occasion. Here, the Traditional Batting Average exceeds even the maximum score and it will become ∞ irrespective of the runs scored if the batter does not get out in any of the innings.

Therefore, the key question becomes: how many runs could a batter potentially score if they were allowed to complete their innings under ideal circumstances? From statistical perspective, the expected runs from 'not out' innings can be estimated by appropriate weights based on the player's batting position i.e. the point in the innings at which they enter to bat.

Based on various studies, position plays a crucial role in scoring capability of a batter. The top three batter get enough opportunity or balls to score more runs according to their scoring ability. Middle-order batsmen often remain "not out" due to match situations like chasing a target, innings ending, or interruptions such as rain or injury. This limits their chance to score freely, making their batting averages less reflective of true ability compared to openers.

A batter often plays at different positions during their career, so expected runs should be computed separately for each position by assigning weights accordingly. Since a batter's score ability is closely associated to their strike rate, the number of balls faced becomes a key factor. For any fixed position, the adjusted score of an incomplete innings is estimated by multiplying by actual 'not out' runs by the ratio of the average balls faced in complete innings at that position to the balls faced in the incomplete innings.

Notations:

Suppose a batter has played m complete innings and n incomplete innings over his/her career in One-Day International. The following variables are defined as

- $X_{(p)i}$: the runs scored in the i^{th} **complete** innings at p^{th} position of batting.
- $X^*_{(p)j}$: the runs scored in the j^{th} **incomplete** innings at p^{th} position of batting.
- $Y_{(p)k}$: the runs scored in the k^{th} **combined** innings at p^{th} position of batting.
- $Z_{(p)j}$: the expected or updated runs scored in the j^{th} **incomplete** innings at p^{th} position of batting.
- $B_{(p)i}$: the the number of ball faced in the i^{th} **complete** innings at p^{th} position of batting.
- $B^*_{(p)j}$: the the number of ball faced in the j^{th} **incomplete** innings at p^{th} position of batting.
- $m_{(p)}$: the number of **complete** innings at p^{th} position of batting, $\sum_{p=1}^{11} m_{(p)} = m$.
- $n_{(p)}$: the number of **incomplete** innings at p^{th} position of batting, $\sum_{p=1}^{11} n_{(p)} = n$.
- $N = m + n$: the total number of combined innings.

$\forall\ i = 1, 2, \dots, m_{(p)},\ j = 1, 2, \dots, n_{(p)}, k = 1, 2, \dots, \overline{m_{(p)} + n_{(p)}}, p = 1, 2, \dots, 11$

Mathematical Formulation of Batting Average

In this context, the **Traditional Batting Average** is defined by,

$$\textbf{TBA} = \frac{1}{m}\left[\sum_{p=1}^{11}\left\{\sum_{i=1}^{m_p} X_{(p)i} + \sum_{j=1}^{n_p} X^*_{(p)j}\right\}\right]$$

For fixed position p,

$$Z_{(p)j} = \begin{cases} X^*_{(p)j}\dfrac{\frac{1}{m_p}\sum_{i=1}^{m_p} B_{(p)i}}{B^*_{(p)j}} & \text{if} \quad \frac{1}{m_p}\sum_{i=1}^{m_p} B_{(p)i} > B^*_{(p)j} \\ X^*_{(p)j} & \text{if} \quad \frac{1}{m_p}\sum_{i=1}^{m_p} B_{(p)i} \leq B^*_{(p)j} \end{cases}$$

The rationale is that a batter's expected score should always be higher than their actual '*not out*' runs, as they have already achieved those runs and could have potentially scored more if the innings had continued.

After successfully updating the 'not out' scores by the aforesaid algorithm, the new measure of the average is the arithmetic mean of the combined runs for the position 1, 2, 3 and the updated runs and the runs of complete innings for the position 4 to 11. Mathematically,

$$\textbf{PBA} = \frac{1}{N}\left[\sum_{p=1}^{3}\sum_{k=1}^{m_3+n_3} Y_{(p)k} + \sum_{p=4}^{11}\left\{\sum_{i=1}^{m_p} X_{(p)i} + \sum_{j=1}^{n_p} Z_{(p)j}\right\}\right]$$

The algorithm was applied to a few pre-selected batters from around the world, and the results are presented in the table below.

Table 1: AM = Arithmetic Mean, TBA = Traditional Batting Average, PBA = Proposed Batting Average

Batter's Name	Country	AM	TBA	PBA	Median Runs
Sourav Ganguly	India	37.88	40.73	37.91	26.0
Sachin Tendulkar	India	40.77	44.72	40.85	28.5
MS Dhoni	India	36.27	50.58	39.42	31.0
Virat Kohli	India	49.65	59.34	49.73	37.0
Yuvraj Singh	India	31.3	36.56	32.27	22.0
Kevin Pietersen	England	35.52	40.73	35.77	27.0
Paul Collingwood	England	28.13	35.36	29.68	19.0
Eion Morgan	England	33.64	39.19	34.46	24.0
Kane Williamson	New Zealand	42.87	47.48	43.0	33.5

Batter's Name	Country	AM	TBA	PBA	Median Runs
Ross Taylor	New Zealand	39.69	48.44	40.78	30.5
Brendon McCullam	New Zealand	26.68	30.41	27.5	18.0
Craig McMilan	New Zealand	25.72	28.19	26.02	18.0
AB de Villiars	South Africa	43.93	53.5	44.38	34.5
Jacque Kallis	South Africa	36.88	44.19	37.35	28.5
JP Duminy	South Africa	28.59	36.81	30.38	23.0
Michael Bevan	Australia	35.27	53.17	37.38	32.0
Ricky Ponting	Australia	37.55	41.91	37.64	30.0
Michael Clarke	Australia	35.79	44.59	37.05	33.0
Mike Hussey	Australia	34.66	48.16	38.55	30.0

Test for Location of the Actual Score and the Updated Score

As there is no Gaussianity assumptions of the runs, the **paired sample Sign test** is conducted in order to check if there is any significance difference between the actual 'not out' scores and expected or updated 'not out' scores.

Interesting Facts

The analysis revealed several interesting insights, which are outlined below.

- Shoaib Malik is among the rare batters to have played in every position from 1 to 10.
- M.S. Dhoni holds the record for the most '*not out*' innings in ODI history.
- Misbah-ul-Haq is the only player to maintain a 40+ batting average in ODIs without scoring a century.