JOURNAL
OF
APPLIED
QUANTITATIVE
METHODS

ON FINDING THE MOST COMPATIBLE BATTING AVERAGE

Prodip Kumar GAUR[1]
Post Graduate Student,
Department of Statistics, Assam University, Silchar, India

E-mail: prodip.786@rediffmail.com

Dibyojyoti BHATTACHARJEE[2]
Professor, Department of Statistics,
Assam University, Silchar, India

E-mail: djb.stat@gmail.com

Abstract
Batting average is the most commonly used measure of batting performance in cricket. It is defined as the total number of runs scored by the batsman divided by the number of innings in which the batsman was dismissed. Generally, the innings of a batsman comes to an end due to his dismissal, yet there are some cases in which the batsman may not get dismissed due to sudden termination of the batting innings of the team. The sudden termination may take place due to bad weather or victory or injury of the batsman or for running short of partners etc. In case, there are several not out innings in the career of a batsman, the batting average may get overestimated. To overcome this problem of over estimation, several authors proposed different modifications to the existing formula of batting average or defined new measures. Though each method expressed its advantages over the existing batting average, yet none of them are universally accepted as the most efficient replacement of the existing formula. This paper makes an attempt to study the existing solutions to the problem and then to evaluate the best or at least the most compatible alternative. For the purpose of quantification, data from the ICC Cricket World Cup played in Australia and New Zealand in 2015 is considered.

Keywords: *Data Mining in Sports; Performance Measurement; Cricket, Batting Average*

1. Introduction

Cricket is an outdoor game played with bat and ball in a specially prepared area in the center of circular field called a pitch. The game is played under certain rules and regulations between two teams of eleven players each. The teams take turn at batting and fielding. Each of such turn is called an innings. The aim of the fielding team is to dismiss all the batsmen of the batting team and/or to restrict the flow of runs. Presently, there are three versions of cricket being played at the international level: test cricket, one-day international cricket (ODI) and Twenty20 cricket (Saikia, Bhattacharjee & Radhakrishnan, 2016). While test match is an unlimited over game, ODI and Twenty20 are restricted over versions of cricket.

The ODI matches are of 50 overs per innings; and Twenty20, as the name indicates are of 20 overs duration only. ODI and Twenty20 format are called limited over format of cricket. In limited over cricket, the team which bats first sets a target for the opponent team to attain in the second innings.

In cricket each batsman try to score as much runs he can against the bowling attack of the fielding team. The bowlers on the other hand, with the help of the fielders try to restrict the batsman from scoring runs. Though both bowling and batting are the prime skills of the game of cricket yet in this work, we shall concentrate on a very common measure of batting performance viz. the batting average. The batting average is an index of the batting ability of a cricketer.

The batting average of a batsman in *'n'* innings (say) is defined as the total runs scored by the batsman in those innings divided by the number of complete innings. The phrase 'complete innings' means the innings in which the batsman was dismissed. If the batsman gets dismissed in all his innings then the batting average is as good as the arithmetic mean of the runs scored by the batsman in those *'n'* innings. However, if the batsman remains not out in some of the innings (antonymous to 'dismissed'), then the numerator remains same i.e. the runs scored by the batsman in those *'n'* innings but the denominator is only the innings in which the batsman was dismissed i.e. less than *n*. Thus, if there is at least one not out innings in the collection of *'n'* innings then the denominator is less than the number of terms in the numerator. This may overestimate the actual batting performance of a batsman, if measured through batting average. A hypothetical situation, in which the batsman is not dismissed in any of his innings, the batting average remains undefined. Many authors addressed this problem and defined different measures to compliment the issues concerning the batting average. Van Staden et al. (2009) gives a summary of all these methods. In this work, we try to explore these options and try to find out the best or the most compatible of the options.

Section 2 of the paper reviews the different types of performance measures in cricket and provides a brief introduction to all the extended batting averages defined by the different authors. The next section of the paper provides in details different formulae of all the extended batting averages. Section 4 is the methodology of the paper where the data source, process of comparison and relevant statistical methods are discussed. The result of the calculations is discussed in Section 5 and the last section concludes the work with some directions for future work.

2. Literature Review

Cricket is a data-rich sport. So different quantitative works by researches based on data generated from cricket are frequently encountered. Out of which a significant amount of work is concentrated towards performance measurement in cricket. Batting and bowling are two prime skills of the game. Thus, different traditional measures are used in cricket to quantify batting and bowling performance of cricketers. While batting average and strike rate are two very commonly used measures of batting performance, bowling average, bowling strike rate and economy rate are commonly used measured of bowling performance. In addition to these measures, several other authors have defined other innovative measures of quantifying batting and bowling skills of cricketers. Mention can be made of the Combine Bowling Rate by Lemmer (2002) and other measures of bowling performance by Beaudoin and Swartz (2003), Kimber and Hansford (1993) and Van Staden (2009). In case of innovative measures concerning batting performance mention can be made of Lemmer (2004), Barr and Kantor (2004), Croucher (2000), Basevi and Binoy (2007), Kimber and Hansford (1993). Brettenny (2010) reviews the different batting and bowling performance measures proposed by different authors.

Out of the traditional measures of quantifying batting performance in cricket the batting average is the most commonly used in all formats. The formula for which is given by,

$$AV = \frac{\text{Number of runs scored}}{\text{Number of complete innings}} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i + \sum_{i=n+1}^{n+m} x_i^*\right) \quad (1)$$

Where $x_i; i = 1,2, \dots n$ denote the runs scored by a batsman in *n* completed innings and $x_i^*; i = n+1, n+2, \dots n+m$ denote the runs scored by a batsman in *m* not-out innings. The disadvantage of using this formula is that it can overestimate the batsman's batting average. Historically, the principle criterion used for comparing batsmen in the game of cricket has been the batting average, but unfortunately, when a batsman has a high proportion of not-out innings, the batsman's batting average will be inflated (Van Staden et al. 2009). The problem of the study looks beyond the works done by Van Staden et al. (2009).

To address this issue, several authors have suggested changes in the formula of batting averages with techniques ranging from the concept of survival analysis to Bayesian estimation. Some of them are Danaher (1989), Lemmer (2008a), Damodaran (2006), Maini and Narayanan (2007) etc. Though each of the methods is developed based on correct statistical logic yet there is no universal acceptance of any of these methods, as a solution to the problem of over estimation existing in (1).

Van Staden et al. (2009) analyzed and compared different methods which are designed to deal with the problem of inflated batting average due to the presence of a high proportion of not-out innings. From the work of Van Staden et al. (2009), one finds that none of the methods clearly outperforms all the other methods. His work only made an empirical comparison of ten different methods but cannot reach to a meaningful conclusion viz. the best method of computing the batting average. This provided us with a motivation to take up the problem.

3. Description of the different Batting Averages

The simplest solution for dealing with the problem of inflated batting average is to use the "real" AV instead of the conventional AV by dividing the number of runs scored in all innings by total number of innings,

$$AV_{real} = \frac{\text{Total runs scored}}{\text{Number of innings}} = \frac{1}{n+m}\left(\sum_{i=1}^{n} x_i + \sum_{i=n+1}^{n+m} x_i^*\right) \quad (2)$$

With AV_{real} the distinction between completed and not out innings is ignored, and, by doing so, the occurrence of inflated averages is completely eliminated (Howells, 2001).

Danaher (1989), proposed the product limit estimator (PLE) to estimate the batting average. The PLE is a non-parametric estimator originally designed by Kaplan and Meier (1958) for the use in life insurance and the actuarial field in general.
With the PLE, all not out batting scores are censored. Then,

$$PLE = \sum_{i=1}^{n} \Delta y_{i:n} \prod_{j=0}^{i-1} \left(1 - \frac{d_j}{c_j}\right) \tag{3}$$

Where $y_{i:n}; i = 1,2, \ldots n$ denote the ranked distinct uncensored scores, $y_{0:n} = 0, \Delta y_{i:n} = y_{i:n} - y_{(i-1):n,} d_j$ is the number of uncensored scores equal to $y_{i:n}$ and c_j the number of censored and uncensored scores greater or equal to $y_{j:n}$. To ensure that the PLE is finite, the maximum score is uncensored, even if it is a not-out score.

Unfortunately the calculation of the PLE is extremely complex, so it is unlikely that the cricketing world shall favour it. Also, after each extra innings of a batsman, the PLE has to be recalculated completely. Furthermore, as pointed out by Danaher (1989), the PLE is insensitive when many of the high scores are not-out scores and hence censored.

Generally a batsman will always have PLE ≤ AV. However, it is interesting to note that the value of the PLE can be greater than that of AV. This can happen when a batsman's highest score is an outlier, that is, when the highest score is much larger than the second highest score and, in effect, the rest of the batsman's scores (Danaher 1989).
Lemmer (2008a) considered innovative estimators of the type

$$e_g = \frac{1}{n+m}\left(\sum_{i=1}^{n} x_i + f_g \sum_{i=n+1}^{n+m} x_i^*\right) \tag{4}$$

Where, the factor f_g is used to adjust the not out scores to obtain completed scores. The simplest estimator of this type is e_2 with, $f_2 = 2$, so that not out batting scores are doubled (Lemmer, 2008a).

$$e_2 = \frac{1}{n+m}\left(\sum_{i=1}^{n} x_i + 2\sum_{i=n+1}^{n+m} x_i^*\right) \tag{5}$$

The justification for the choice of $f_2 = 2$ is that, if a batsman had a not out score and assuming that the batsman would be allowed to continue till he gets dismissed, then, on average, he could have been expected to double his score.

Lemmer (2008a), also considered many other possible factors, and found that e_6 with $f_6 = 2.2 - 0.01\bar{x}^*$, where $\bar{x}^* = \frac{1}{m}\sum_{i=n+1}^{n+m} x_i^*$ is the average of the not out batting scores.

Thus we define e_6 as,

$$e_6 = \frac{1}{n+m}\left(\sum_{i=1}^{n} x_i + \left(2.2 - 0.01\bar{x}*\right)\right)\sum_{i=n+1}^{n+m} x_i * \qquad (6)$$

Lemmer (2008a) showed that e_2 and e_6 are closely related. But the calculation of e_6 is more complicated than that of e_2, accordingly he suggested that e_2 can be used for ease in calculations without much compromise with accuracy. Lemmer (2008b) recommended,

$$e_{26} = \frac{1}{2}\left(e_2 + e_6\right) \qquad (7)$$

Van Staden et al. (2009), defined another simpler measure like that of e_2 with an interesting modification. According to that method, the runs scored in the not out innings is either doubled or restricted to the highest score achieved by the batsman in the past tournament or career innings. Out of the two options, the minimum shall be considered for a not out innings. It is denoted by e_2^r.

Damodaran (2006), utilized a Bayesian approach to replace not out scores with conditional average scores. Consider the series of innings *t* = *1, 2,…,n+m*, if the score in innings *t* is a complete score, x_t then we take $z_t = x_t$. If the score is a not out score, x_t^*, then this score is replaced by

$$z_t = \text{int}\left\{\frac{\sum_{l=1}^{t-1} z_l I\left(z_l\right)}{\sum_{l=1}^{t-1} I\left(z_l\right)}\right\} \qquad (8)$$

where $Z_1, Z_2, \dots, Z_{t-1}$ are the series of completed and/or adjusted scores up to innings $t-1, I(Z_l) = 0 \; if \; x_i^* \geq Z_l and \; I(Z_l) = 1 \; if \; x_i^* \leq Z_l$

The estimator for the average is then given by,

$$AV_{Bayesian} = \frac{1}{n+m}\sum_{t=1}^{n+m} Z_t \qquad (9)$$

Maini and Narayanan (2007) proposed a method based upon exposure-to-risk. Let

$$\bar{b} = \frac{\text{Number of balls faced}}{\text{Number of innings}} = \frac{1}{n+m}\left(\sum_{i=1}^{n} b_i + \sum_{i=n+1}^{n+m} b_i^*\right) \qquad (10)$$

Be the average number of balls faced by a batsman in his $m+n$ innings and let $r_1, r_2, \dots, r_n$ and $r_{n+1}^*, r_{n+2}^*, \dots, r_{n+m}^*$ denote the batsman's exposure in *n* completed innings and *m* not out innings respectively. If the score in innings i is a completed score, $r_i = 1$. In effect the exposure is one for all completed innings. If the score is a not out score and $b_i^* < \bar{b},$ then, $r_i^* = \frac{b_i^*}{\bar{b}}$, else $r_i^* = 1$. The average is then calculated by

$$AV_{\exp osure} = \frac{\sum_{i=1}^{n} x_i + \sum_{i=n+1}^{n+m} x_i^*}{\sum_{i=1}^{n} r_i + \sum_{i=n+1}^{n+m} r_i^*} \tag{11}$$

Van Staden et al. (2009) pointed out two issues with the $AV_{exposure}$ - first, the number of balls faced by a batsman in a not-out innings is compared to the average number of balls faced over the whole tournament or career of this batsman. Thus, the exposure calculated for a not-out innings depends on past and future batting performances, which is not logical. Surely only past batting performances should be used. Further, the exposure for each past not-out innings must be recalculated each time the batsman bats again. So an immediate advantage of only using past batting performances will be that the exposure for past not-out innings need not be recalculated after each additional innings. The second concern has to do with the calculation of the average number of balls faced. Accordingly, Van Staden et al. (2009) suggested that a batsman should benefit from surviving the opposition's bowling attack by comparing the number of balls faced in a not-out innings to the survival rate instead of the average number of balls. Applying both the adjustments to the exposure-to-risk method, if a batting score is a not-out score and $b_i^* < SV_i$ where SV_i , is the survival rate for the batsman for all innings up to and including innings *i*, then $r_i^* = \frac{b_i^*}{SV_i}$, else $r_i^* = 1$. Accordingly, Van Staden et al. (2009) denoted the average based upon our adjusted exposure-to-risk method by $AV_{survival}$ to distinguish it from exposure AV_{exposure}.

4. Methodology

To compare the different averages discussed so far we need to apply it to some live data and compute the different averages. The conformity between the different methods shall be checked by the Kendall's coefficient of concordance and then in case of non-conformity sensitivity analysis shall be performed to find out the average that has maximum compatibility with the other averages. Detailed explanation of data source and the methodologies are explained in the subsequent subsections.

4.1. Data Source and Training Sample

For computing the batting averages using different methods and then for further relevant computations to reach the objective of the study we need a real data set. For the dataset, the matches played in the 2015 World Cup in Australia and New Zealand, is considered. The world cup of 50-overs a side saw 49 matches in the tournament. The necessary data from those matches are collected from the website www.espncricinfo.com. For the purpose of the study, the batsmen who satisfy the following criteria are considered in the trial sample for the computation of batting averages using the different methods:

- The batsman who has played at least 5 innings in the entire tournament
- The batsman who was not-out in at least one innings in the entire tournament
- The batsman who has faced at least 200 balls in the entire tournament

4.2. Computation of Averages

Following the restrictions as in the previous section, 20 batsmen qualified for the training sample, details of which are provided in Appendix I. Based on the different methods discussed above the computation are done and the averages along with the ranks are summarized in Table 1.

4.3. Kendall's Coefficient of Concordance

Since the formula for computation, varies from each other so it is obvious that the computed average using different methods will give different values even for the same batsman. However, the ranks of the batsmen shall not be much variant across the different methods, if based on the same data. Thus, once the averages of the batsmen are obtained using different methods, the batsmen shall be ranked based on each of the methods. Then considering each method of average as one of the rater, Kendall's coefficient of concordance shall be computed. Kendall's coefficient of concordance is a measure of agreement among raters and is defined as follows.

Assume there are *m* raters (here 10 different method of averages) rating *k* subjects (here 20 different batsmen) in rank order from 1 to *k*. Let $r_{ij} =$ the rating rater *j* gives to subject *i*. For each subject *i*, let $R_i = \sum_{j=1}^{m} r_{ij}$ let $\overline{R}$ be the mean of the R_i, and let *R* be the squared deviation, i.e.

$$R = \sum_{i=1}^{k} \left(R_i - \overline{R}\right)^2 \quad (12)$$

Now we define Kendall's *W* by

$$W = \frac{12R}{m^2\left(k^3 - k\right)} \quad (13)$$

It is also to be noted that the value of W always lies between 0 and 1 *i.e.* $0 \le W \le 1$. It is given that by the first property of Kendall's coefficient of concordance when $k \ge 5$ or *m* > 15, $m(k-1)W \sim \chi^2_{k-1}$. This rule can be used to test the null hypothesis that all the raters (averages) have ranked the subjects (batsmen) in a uniform manner.

4.4. Pareto Ordering for Compatibility

If the null hypothesis mentioned in the previous sub-section is rejected for the exercise on the current data set, then it means that the different methods of averaging have not ranked the batsmen in a uniform way but differently. In such a case, we can take the help of Pareto ordering to determine that average (set of ranks) which has the maximum compatibility with the other averages (set of rankings). Chakrabarty and Bhattacharjee (2012), can be consulted for detailed discussion of the Pareto ordering method. In brief, its working can be explained by the following way,

Let, the subscript *i* is an index attributed to identify the batsman. Since, there are 20 batsmen in the training sample so $i = 1, 2, \ldots, 20$ and the subscript *j* (or *k*) is an index attributed to the method of averaging. As, there are 10 method of averages discussed in the paper so *j* (or *k*) $= 1, 2, \ldots, 10$. Next, we define,

R_i^j = Rank of the i^{th} batsman in the j^{th} method of computing average

d_i^{jk} = Square of difference between ranks of the i^{th} batsman for the j^{th} and k^{th} method of computing average = $\left(R_i^j - R_i^k\right)^2$

D^j = Sum of square of distance between ranks of the j^{th} method of averaging with all other methods across all batsmen = $\sum_{k \neq j=1}^{10} \sum_{i=1}^{20} d_i^{jk}$ (14)

So, the compatibility score corresponding to the j^{th} method of averaging is given by D^j as defined in (14). Lesser the compatibility score of a given method of average more is the compatibility of that average with a set of similar other method of averages.

5. Results and Discussion

Considering the data restriction mentioned above 20 batsmen got selected in the training sample. The batting average of all of them is calculated using the different method of averages and are placed in Table 1 below.

Table 1. Averages of the batsmen in the training sample under the different methods

Player name	AV	AV_{real}	PLE	e_2	e_6	e_{26}	e_2^r	$AV_{Bayesian}$	$AV_{exposure}$	$AV_{survival}$
SPD Smith	67(6)	57.43(5)	61.46(3)	65.43(8)	62.55(2)	63.99(5)	57.43(5)	57.43(5)	57.43(6)	59.04(5)
David Warner	49.29(13)	43.13(11)	48.058(9)	45.75(16)	45.72(10)	45.74(14)	43.13(11)	43.13(11)	48.14(9)	48.69(9)
Mahmudullah	73(3)	60.83(3)	56.67(7)	82.17(4)	59.13(4)	70.65(3)	60.83(3)	60.83(3)	60.83(3)	61.98(3)
IR Bell	52.4(11)	43.67(10)	44.75(13)	52.33(11)	49.56(7)	50.95(10)	43.67(10)	43.67(10)	43.73(11)	46.28(12)
MS Dhoni	59.25(9)	39.5(13)	44.5(14)	61.17(9)	34.63(14)	47.9(11)	39.5(13)	39.5(13)	39.5(13)	45.69(13)
Suresh Raina	56.8(10)	47.33(8)	41.33(15)	65.67(7)	49.17(8)	57.42(7)	47.33(9)	47.33(9)	47.33(10)	46.6(11)
Virat Kohli	50.83(12)	38.13(15)	46.42(11)	47.75(15)	37.23(12)	42.29(15)	38.13(15)	38.13(15)	39.23(14)	47.24(10)
Rohit Sharma	47.14(14)	41.25(12)	46.69(10)	48.38(14)	45.74(19)	47.06(12)	41.25(12)	41.25(12)	41.25(12)	42.97(14)
Ajinke Rahane	34.67(17)	29.71(17)	33.79(16)	33.43(17)	33.82(15)	34.12(17)	29.71(17)	29.71(17)	30.69(18)	32.22(18)
MJ Guptill	68.38(4)	60.78(4)	46.11(12)	87.11(3)	29.97(17)	58.84(6)	60.78(4)	60.78(4)	60.78(4)	55.78(7)
GD Elliot	44.29(15)	38.75(14)	60.98(4)	49.25(13)	42.53(11)	45.89(13)	38.75(14)	38.75(14)	38.75(15)	35.54(15)
KS Williamson	33.43(18)	26(19)	30.59(18)	32(18)	25.79(19)	28.9(19)	26(19)	26(19)	27.02(20)	31.59(19)
CJ Anderson	33(19)	28.88(18)	32.67(17)	29.75(19)	29.86(18)	29.81(18)	28.88(18)	28.88(18)	31.64(17)	32.34(17)
LRPL Taylor	31.57(20)	24.56(20)	29.7(19)	27.78(20)	24.64(20)	26.21(20)	24.56(20)	24.56(20)	27.2(19)	27.56(20)
AB de Villiers	96.4(2)	68.86(2)	75.67(2)	101.29(2)	53.7(6)	77.49(2)	78.14(1)	78.321(1)	69.42(2)	78.44(2)
DA Miller	64.8(7)	46.29(9)	50.95(8)	72.57(5)	36.82(13)	54.7(9)	52.86(8)	53.07(8)	48.42(8)	51.44(8)
F du Plessis	63.33(8)	54.29(7)	59.83(5)	57.29(10)	57.26(5)	57.27(8)	54.29(7)	54.29(7)	58.61(5)	59.72(4)
KC Sangakarra	108.2(1)	77.29(1)	101.63(1)	109(1)	62.86(1)	85.93(1)	77.29(2)	77.29(2)	77.29(1)	90.13(1)
MN Sammuels	38.33(16)	32.86(16)	19.29(20)	51.86(12)	30.39(16)	41.12(16)	32.86(16)	32.86(16)	32.86(16)	34.71(16)
SC Williams	67.8(5)	56.5(6)	59.82(6)	69.17(6)	62.07(3)	65.62(4)	56.5(6)	56.5(6)	56.5(7)	58.11(6)

In Table 1, the name of the batsman appears along the row heads and the method of average along the column heads. The number in the cell indicates the average of the batsman appearing in the row head using the method of average indicated by the column head. The numbers in parenthesis in each of the cells shows the rank of the cricketer depending on the method of averaging as given in the column head. It may be seen that KC Sangakara is ranked first in most of the method eight out of ten and LRPL Taylor is ranked last (20th) in eight out of ten methods.

Now, Kendall's coefficient of concordance is computed for the data set, with an aim to test the null hypothesis W = 0, which is an indication that there is agreement among the methods. The computation for the data set under consideration provides, W = 0.9034 with the *p*-value of the corresponding χ^2 statistic as 0.000 indicating that there is a clear disagreement between the different method of averages.

Next we perform Pareto ordering and compute the values of D^i following (14) above. Table 2 provides the compatibility score of the different batting averages. Minimum the value of D^i more compatible is the method of averaging. The table shows that, e_2^r and $AV_{Bayesian}$ has maximum compatibility with the other methods of averaging. AV_{real} acquires the next position in compatibility with very close compatibility score with e_2^r and $AV_{Bayesian}$. However, considering the simplicity of AV_{real}, one may consider it as the best method of computing batting average.

Table 2. Compatibility Score (D^i) of the different method of computing the batting average

AV	AV_{real}	*PLE*	e_2	e_6	e_{26}	e_2^r	$AV_{Bayesian}$	$AV_{exposure}$	$AV_{survival}$
1060	870	2278	1750	2988	952	868	868	1008	1090

6. Direction of Future Work and Conclusion

Of the different formats of cricket, this exercise is performed here only on one day international (ODI) cricket that too for a tournament only. In this exercise, we found that e_2^r and $AV_{Bayesian}$ has maximum compatibility compared to the other method of averages. In order to generalize the result, several such exercises shall be performed over different sets of one day international matches. The exercise can be extended to other formats of cricket like, Twenty20 and test cricket as well. This can enable the researcher to understand how the batting average shall be best defined depending on the format of cricket. However, what is statistically correct may not be accepted to cricket analysts and fans. The batting average needs to be well and simply defined, so that any cricket fan can easily compute the average. In this regard, in this exercise, AV_{real} even being in the second position shall earn more acceptance than others.

References

1. Barr, G. D. I and Kantor, B. S. **A criterion for comparing and selecting Batsmen in limited overs Cricket**, Journal of the Operational Research Society, Vol. 55, 2004, pp. 1266-1274
2. Basevi, T. and Binoy, G. **The worlds best Twenty20 players,** http://content-rsa.cricinfo.com/columns/content/story/311962html, 2007

3. Beaudoin, D. and Swartz, T. **The best Batsmen and Bowlers in One-day Cricket**, South African Statistical Journal, Vol. 37, 2003, pp. 203-222
4. Brettenny, W. **Integer optimisation for the selection of a fantasy league cricket team**, Unpublished Ph.D Dissertation, 2010
5. Chakrabarty, N. and Bhattacharjee, D. **Aggregation and weightage issues concerning composite index development: experience with digital divide in Asian countries**, Asia Pacific Journal of Library and Information Science, Vol. 2, 2012, pp. 24–37
6. Croucher, J. S. **Player rating in one-day cricket,** Proceedings of the Fifth Australian Conference on Mathematics and Computers. NSW, 2000, pp. 95-106.
7. Damodaran, U. **Stochastic dominance and analysis of ODI batting performance:the indian cricket team, 1989-2005**, Journal of Sports Science and Medicine,Vol. 5, No. 4, 2006, pp. 503-508
8. Danaher, P.J. **Estimating a cricketer's batting average using the product limit estimator**, New Zealand Statistician,Vol. 24, No. 1, 1989, pp. 2-5
9. Kaplan, E.L. and Meier, P. **Nonparametric estimation from incomplete observations,** Journal of American Statistical Association,Vol. 53, No. 282, 1958, pp. 457-481
10. Kimber, A. **A Graphical display for Comparing Bowlers in Cricket**, An International Journal for Teachers, Vol. 15, No. 3, 1993, pp. 84-86
11. Kimber, A. C. and Hansford, A.R. **A Statistical Analysis of Batting in Cricket**, Journal of the Royal Statistical Society , Vol. 156, No. 3, 1993, pp. 443-455
12. Lemmer, H H. **A Measure for the Batting Performance of Cricket Players**, South African Journal for Research in Sport, Physical Education and Recreation, Vol. 26, No. 1, 2004, pp. 55-64
13. Lemmer, H H. **The Combined Bowling Rate as a Measure of Bowling Performance in Cricket**, South African Journal of Research in Sport, Physical Education and Recreation, Vol. 24, No. 2, 2002, pp. 37-44
14. Lemmer, H. H. **An analysis of players' performances in the first cricket Twenty20 WOrld Cup series**, South African Journal for Research in Sport, Physical Education and Recreation,Vol. 30, No. 2, 2008b, pp. 71-77
15. Lemmer, H. H. **Measures of batting performance in a short series of cricket matches**, South African Statistical Journal, Vol. 42, No. 1, 2008a, pp. 65-87
16. Maini, S. and Narayanan, S. **The flaw in batting averages**, The Actuary, May-07, 2007, pp. 30-31
17. Saikia, H., Bhattacharjee, D. and Radhakrishnan, U. K. **A New Model for Player Selection in Cricket**, International Journal of Performance Analysis in Sports, Vol. 16, 2016, pp. 373-388
18. van Staden, P.J. **Comparison of cricketers bowling and batting performances using graphical displays**, Current Science, Vol. 96, No. 6, 2009, pp. 764-766
19. Van Staden, P.J., Meiring A.T., Steyn J.A., Fabris-Rotelli I.N. **Meaningful batting averages in cricket**, Annual Congress of the South African Statistical Association, November 2010, Potchefstroom, South Africa. Unpublished conference paper

Appendix 1. Innings wise performance of the batsmen in the training sample

Innings→ Batsman	1	2	3	4	5	6	7	8	9
SPD Smith	5(9)	4(11)	95(98)	72(88)	65(69)	105 (93)	56* (71)		
David Warner	22(18)	34 (42)	178 (133)	9(12)	21* (6)	24 (23)	12(7)	45 (46)	
Mahmadullah	23 (46)	28 (46)	62 (62)	103 (138)	128* (123)	21 (31)			
I R Bell	36(45)	8(17)	54(85)	49(54)	63(82)	52*(56)			
M S Dhoni	18(13)	18 (11)	45* (56)	85* (76)	6(11)	65(65)			
S R Raina	74(56)	6(5)	22(25)	110* (104)	65(57)	7(11)			
Virat Kohli	107 (126)	46 (60)	33* (41)	33(36)	44* (42)	38(48)	3(8)	1(13)	
Rohit Sharma	15(20)	0(6)	57*(55)	7(18)	64(66)	16(21)	137 (126)	34(48)	
Ajinka Ra-hane	0(1)	79 (60)	14(34)	33* (28)	19(24)	19(37)	44(68)		
MJ Guptill	49(62)	17 (14)	22(22)	11(14)	57(76)	105 (100)	237* (163)	34(38)	15 (34)
GD Elliot	29 (34)	29(31)	0(1)	19(28)	39(34)	27(11)	84(73)	83(82)	
KS Williamson	57(65)	38 (45)	9(22)	45(42)	33(45)	1(2)	33(35)	6(12)	12 (32)
CJ Anderson	75(46)	11 (16)	26 (42)	7(8)	39(26)	15(16)	58(57)	0(2)	
LRPL Taylor	14(28)	9(14)	5(5)	1(2)	24(41)	56(97)	42(61)	30 (39)	40 (72)
AB devilliers	25(36)	30 (38)	162* (66)	24(9)	77(58)	99(82)	65* (45)		
DA Miller	138*(2)	22(23)	20(16)	46*(23)	0(13)	49(48)	49(18)		
F duplessis	24(32)	55 (71)	62(70)	109 (109)	27(29)	21* (31)	82 (107)		
K Sangakarra	39(38)	7* (13)	105* (76)	117 (86)	104 (107)	124 (95)	45(96)		
Marlon Sam-muels	21(41)	38 (52)	133* (156)	0(9)	2(7)	9(18)	27(15)		
SC Williams	8(13)	76* (65)	76(61)	33(32)	96(83)	50(57)			

Source: http://www.espncricinfo.com/icc-cricket-world-cup 2015/engine/series/509587.html? view=records
Note: Not out innings is indicated by *. Figures in bracket indicate the number of balls faced.

---

[1]Prodip Kumar Gaur, is a fresh Post Graduate from the Department of Statistics, Assam University, Silchar, Assam, India. His specialization is in Operations Research and Industrial Statistics.

[2] Professor Dibyojyoti Bhattacharjee has acquired his Post Graduate in Statistics and M. Phil (Statistics) both from the University of Delhi. He completed his Ph.D on Statistical Graphics from Gauhati University.
He worked in institutes like Central Statistical Organization, New Delhi, G C College, Gauhati University, Assam University. Currently he is the Head of the Department of Statistics at Assam University.
He has written about 15 books in different sub-fields of Statistics and published several research papers in different national and international reputed journals.