Ranking the cricket captains using principal component analysis

Article · June 2018

2 authors:

Parag Shah
H L College of Commerce
26 PUBLICATIONS   52 CITATIONS

SEE PROFILE

M. N. Patel
4 PUBLICATIONS   4 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Medical Science View project

Minor Research Project by UGC View project

International Journal of Physiology, Nutrition and Physical Education

Dr. Parag Shah
Department of Statistics, HL
College of Commerce,
Ahmedabad, Gujarat, India

Dr. MN Patel
Department of Statistics,
Gujarat University, Ahmedabad,
Gujarat, India

Ranking the cricket captains using principal component analysis

Dr. Parag Shah and Dr. MN Patel

Abstract
Every cricket match generates an enormous volume of data which is a delight for statisticians and data scientist. Lot of work has been done for performance measurement, decision making, team selection, score prediction and so on. Research has been done in ranking batsman or a bowler or all-rounder based on several batting or bowling measures but very little work has been done in ranking the captains. In this paper, we have ranked captains based on several parameters using Principal Component Analysis. Also, we have included weighted average method to rank captains based on z score of performance of team, individual performance of captain as batsman and bowler.

Keywords: batting performance, bowling performance, team performance, captaincy, winning percentage, principal component analysis, weighted average method

Introduction

Wood (1945) [17] examined the performance of consistency in cricket and applied the geometrical distribution to model cricket scores based on results from test cricket. Kimber and Hansford (1993) [2] proposed a non-parametric approach based on runs scored for assessing batting performance of cricketers. Two comprehensive measure, called as combined bowling rate (CBR) and dynamic bowling rate (DBR) developed by Lemmer in 2002 and 2005 [3, 5] to measure the current bowling performance of the bowlers in one-day and test cricket respectively. Lemmer (2004, 2006, and 2008) [4, 6, 7] developed a classification scheme for batsmen and bowlers using performance data of one-day international (ODI) matches, Test cricket and T 20. The use of stochastic dominance rules demonstrates by Damodaran (2006) [1] to analyze the batting performance of Indian cricketers in ODI cricket. Tan and Ramachandran (2010) [14] ranked the all-rounders in test cricket, both past and present, in accordance with a mathematical formula derived from batting and bowling records. Shah Parag (2017) [13] has defined a new batting and bowling measure. He has defined batting average considering the quality of bowler he is facing and similarly bowling average considering the quality of batsman he is bowling to.

Van Staden (2008) [15] used the term ideal all-rounder, batting all-rounder and bowling all-rounder to characterize the performance of all-rounders. Van Staden (2009) [16] developed a performance measure for cricketers' in Twenty20 cricket considering data from IPL-I. Lenten, Geerlind and Konya (2009) [8] compared a range of cross sectional models to study the factors determining the performance of cricketers in different forms of the game including IPL. Naik & Khattree (1996) [11] provided an example in which principal component analysis is used with sports data. Manage, Ananda B.W. Scariano, Stephen M. (2013) [9] in his paper has used Principal Component Analysis for ranking the players in cricket. More recently Radicchi (2011) [12] applied PageRank algorithm to rank tennis players. Mukherjee (2012) [10] has identified the greatest captain using complex network approach.

In cricket the captain is responsible for the team. Before the game starts the home captain tosses the coin and the touring captain calls heads or tails. The captain chooses the batting order, sets up fielding positions and shoulders the responsibility of on-field decision-making. Thus, the outcome of a match depends on the captain's decisions. In this sense, the success of a team depends on the captain. However, currently there exist no ranking schemes to rank the cricket captains. While the number of wins is a natural measure for success of a team, it does

Correspondence
Dr. Parag Shah
Department of Statistics, HL
College of Commerce,
Ahmedabad, Gujarat, India

not provide a full picture of the 'quality' of win. In this paper, we have considered captain's batting, bowling performance and also team performance and have ranked them using Principal Component Analysis. We have also ranked captains using weighted average method based on z score of performance of team, individual performance of captain as batsman and bowler. A comparison of ranking captains using currently used winning percentage, principal component analysis and weighted average method is also done.

Methodology
Data: Batting, Bowling performance of each Captain and team performance under his captaincy. Data of all Captains in One day International (from 2005 to April, 2018) who have captained the team for atleast 40 matches.

Captain's ranking based on winning percentage
In current scenario performance of captains are evaluating by the number of matches their teams have won under his captainship i.e. only winning percentage are considered. Following table 1 shows the winning percentage of captains.

Rank	Player	Winning %
1	V Kohli (INDIA)	79.16
2	RT Ponting (AUS)	76.14
3	MJ Clarke (AUS)	70.42
4	KC Sangakkara (SL)	65.85
5	GC Smith (Afr/SA)	64.23
6	EJG Morgan (ENG)	62
7	BB McCullum (NZ)	61.86
8	Asghar Stanikzai (AFG)	60.97
9	AB de Villiers (SA)	60.1
10	MS Dhoni (INDIA)	59.57
11	DPMD Jayawardene (SL)	59.09
12	MP Vaughan (ENG)	58.92
13	Mortaza (BDESH)	56.6
14	DL Vettori (NZ)	55.33
15	AN Cook (ENG)	54.47
16	Misbah-ul-Haq (PAK)	53.48
17	KS Williamson (NZ)	52.94
18	SPD Smith (AUS)	52.08
19	AD Mathews (SL)	50.53
20	WTS Porterfield (IRE)	48.35
21	Shakib Al Hasan (BDESH)	46.93
22	AJ Strauss (ENG)	45.08
23	Habibul Bashar (BDESH)	42.02
24	DJG Sammy (WI)	39
25	SO Tikolo (KENYA)	38.02
26	CH Gayle (WI)	36.17
27	JO Holder (WI)	31.13
28	P Utseya (ZIM)	30.14
29	E Chigumbura (ZIM)	29.03

Using the above method, Virat Kohli of India is the number one captain with winning percentage of 79% followed Ricky Pointing of Australia with 76%. Last in the list is Chigumbura of Zimbabwe with winning percentage of 29 percentage.

Principal Component Analysis Method
Principal Component Analysis (PCA) is a nonparametric variable reduction technique well-suited for correlated data that can be effectively used in our context. One objective of principal component analysis is to collapse a set of correlated variables into fewer uncorrelated variables as linear combinations of the original variables.

PCA is particularly useful when data on a number of useful variables has been gathered, and it is plausible that there is some redundancy in those variables. Here, redundancy is taken to mean that our cricket performance variables are correlated with one another because, in some unknown sense, they might be measuring similar player-performance attributes. PCA aims to reduce the observed variables down to a smaller number of principal components, sometimes called auxiliary variables (optimized linear combinations of the original variables), which account for most of the variation occurring in the originally observed variables. These can be utilized to provide summarized measures of performance. For details of PCA, one can refer Hair *et al*. 2007.

This analysis includes the batting measures like Runs, Batting Average (Ave), Batting Strike Rate (SR), Fours, Sixes, Hundreds and Fifty variables, bowling measures like Wickets, Bowling Average, Bowling Strike Rate, team performance like matches played, match won, match lost, discussed in below section, for all captains who have captained for at least 40 ODI matches. This accounts for 29 total captains.
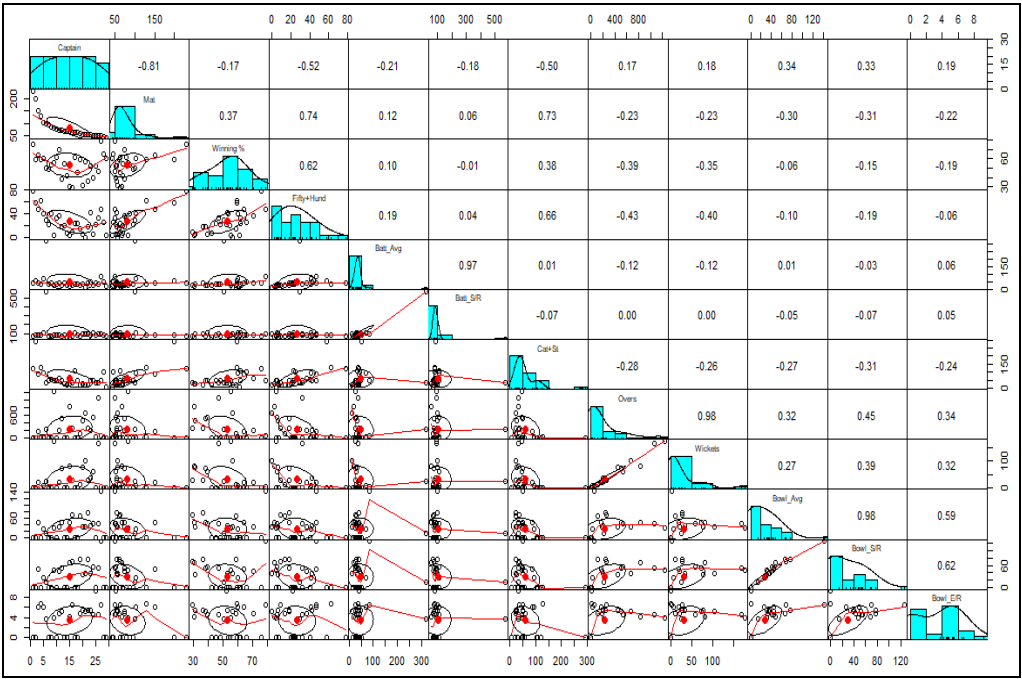
Fig 1: Sample Correlation Matrix for 29 Captains

Table 2: Total variance explained by Components

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	4.097	37.242	37.242
2	2.062	18.747	55.988
3	1.863	16.934	72.922
4	1.369	12.449	85.371
5	.656	5.959	91.331
6	.477	4.340	95.671
7	.281	2.554	98.225
8	.162	1.474	99.699
9	.017	.154	99.853
10	.012	.111	99.964
11	.004	.036	100.000
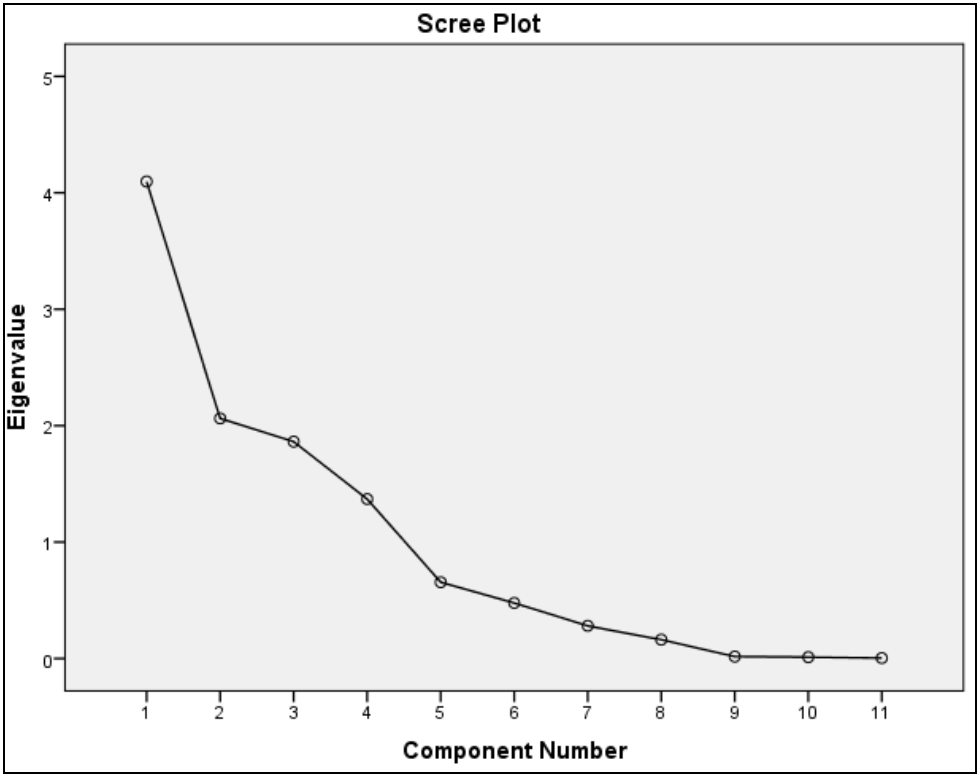
Fig 2: Scree Plot
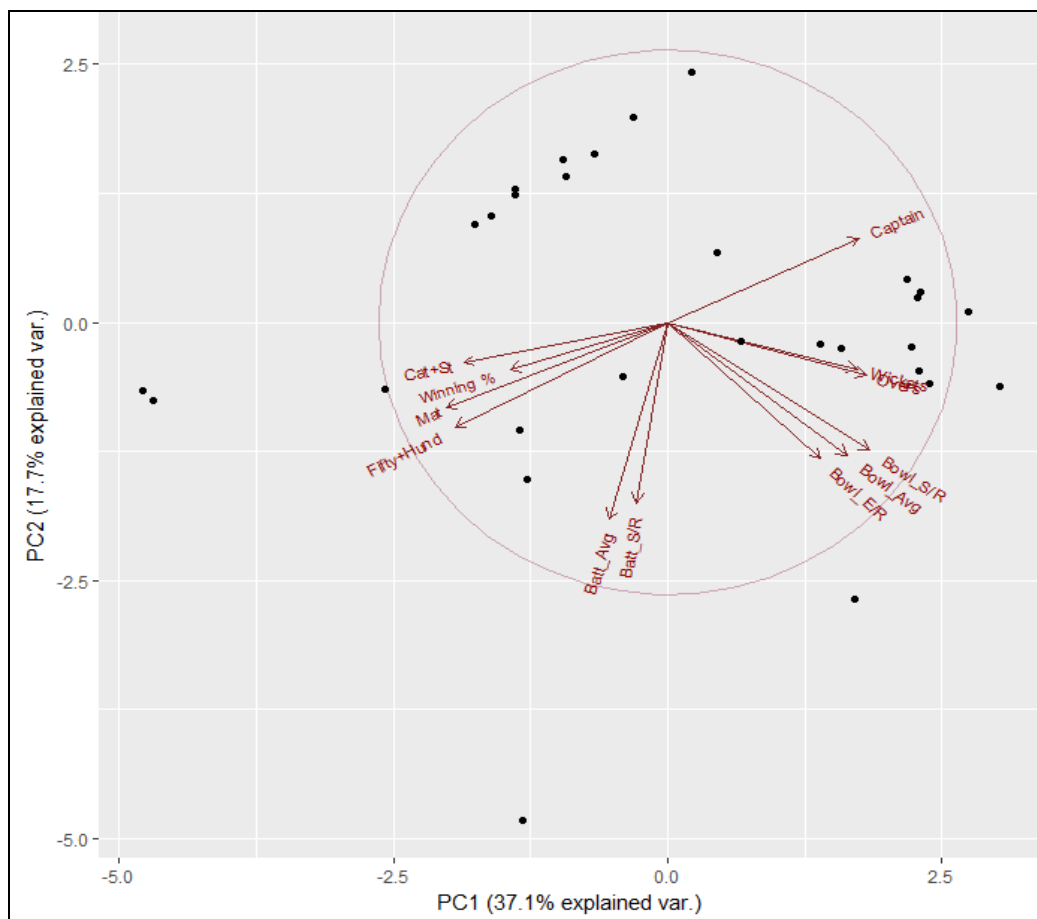
Fig 3: Loading Plot

Table 3: Component Score Coefficient Matrix

Component Score Coefficient Matrix				
	Component			
	1	**2**	**3**	**4**
Matches	.374	-.046	.188	.020
Winning %	.198	.124	-.163	-.020
Fifty + Hundred	.339	.118	-.042	.014
Batt_Avg	.001	.028	-.023	.493
Batt_S/R	-.030	-.033	.051	.505
Catches +Stumping	.354	-.036	.144	-.058
Overs	.091	-.026	.492	.012
Wickets	.099	-.044	.503	.011
Bowl_Avg	.029	.415	-.119	-.023
Bowl_S/R	.034	.383	-.039	-.031
Bowl_E/R	.050	.304	.006	.043

Consequently, the first principal component for Captain is
L1 = 0.374* Matches + 0.198* Winning %
+ 0.339* (Fifty + Hundred) + 0.001* Batt_Avg - 0.030* Batt_S/R
+ 0.354* (Catches + Stumping)
+ 0.091* Overs + 0.099 * Wickets + 0.029 * Bowl_Avg + 0.034* Bowl_S/R
+ 0.050 * Bowl_E/R

We refer to our first principal component as the general – captain – performance – index, which is a type of weighted average of all variables used. Here, the coefficients of the first principal component are mixture of positive and negative, so larger values of L_1 indicate better Captain's performance. This justifies that we should rank (largest to smallest) the captains based on the first principal component.

Table 4: Rank of Captains using Principal Component Analysis score

Rank	Player	PCA
1	MS Dhoni (INDIA)	2.90117
2	RT Ponting (AUS)	2.6454
3	DPMD Jayawardene (SL)	1.42782
4	GC Smith (Afr/SA)	1.31151
5	AB de Villiers (SA)	0.6766
6	V Kohli (INDIA)	0.5646
7	MJ Clarke (AUS)	0.39754
8	DL Vettori (NZ)	0.15717
9	Shakib Al Hasan (BDESH)	0.13584
10	Misbah-ul-Haq (PAK)	0.02903
11	KS Williamson (NZ)	0.00568
12	BB McCullum (NZ)	-0.03282
13	KC Sangakkara (SL)	-0.08187
14	EJG Morgan (ENG)	-0.09218
15	AD Mathews (SL)	-0.21379
16	SO Tikolo (KENYA)	-0.26839
17	WTS Porterfield (IRE)	-0.31966
18	MP Vaughan (ENG)	-0.50974
19	Mortaza (BDESH)	-0.54931
20	SPD Smith (AUS)	-0.59449
21	AN Cook (ENG)	-0.63614
22	AJ Strauss (ENG)	-0.65241
23	P Utseya (ZIM)	-0.71292
24	DJG Sammy (WI)	-0.74829
25	CH Gayle (WI)	-0.77489
26	E Chigumbura (ZIM)	-0.85417
27	JO Holder (WI)	-0.93413
28	Habibul Bashar (BDESH)	-1.10225
29	Asghar Stanikzai (AFG)	-1.17491

Weighted Average Method

In Principal Component Analysis, team's performance and captain's individual performance (batting/bowling) are carrying same weights. One school of thought says that team's performance is more important factor in evaluating captain's efficiency compared to his individual performance. It is the role of the captain to see that team perform well under his leadership and so crucial factor for measuring his captaincy will be team's performance.

In this method we have assigned 60% weightage to the team's performance and 40 % weightage to captain's individual performance. Here we have standardized all the parameters of team performance and combined it to one score called z team. Similarly, we have standardized all the parameters of individual performance (batting and or bowling) and combined it one score z com.

Final Z_Score of a captain is $0.6 \times z$ team $+ 0.4 \times z$ com.

Using this final Z_Score, we have then ranked captains.

Table 5: Rank of Captains using Principal Component Analysis score

Rank	Player	Z_Score
1	V Kohli (INDIA)	4.04
2	AD Mathews (SL)	3.6
3	GC Smith (SA)	2.23
4	RT Ponting (AUS)	2.04
5	DL Vettori (NZ)	2.02
6	MS Dhoni (INDIA)	2.01
7	Shakib Al Hasan (BDESH)	2.01
8	AB de Villiers (SA)	1.12
9	MJ Clarke (AUS)	1.09
10	DPMD Jayawardene (SL)	1.02
11	Mashrafe Mortaza (BDESH)	0.41
12	KS Williamson (NZ)	0.39
13	MP Vaughan (ENG)	0.19
14	SO Tikolo (KENYA)	0.11
15	CH Gayle (WI)	-0.13
16	P Utseya (ZIM)	-0.36
17	DJG Sammy (WI)	-0.37
18	E Chigumbura (ZIM)	-0.67
19	JO Holder (WI)	-1.02
20	SPD Smith (AUS)	-1.12
21	KC Sangakkara (SL)	-1.23
22	BB McCullum (NZ)	-1.24
23	EJG Morgan (ENG)	-1.38
24	Misbah-ul-Haq (PAK)	-1.55
25	WTS Porterfield (IRE)	-2.2
26	AN Cook (ENG)	-2.23
27	AJ Strauss (ENG)	-2.55
28	Asghar Stanikzai (AFG)	-2.88
29	Habibul Bashar (BDESH)	-3.33

A per the weighted average method, Virat Kohli of India tops the captains list followed by Mathews of Sri-lanka. Habibul Bashar of Bangladesh is the bottom of the list.

Table 6: Comparison of ranks of captains using winning percentage, principal component method and weighted average method

Rank	Captains as per Winning %	Captains as per Principal Component Analysis	Captains as per weighted Avg.
1	V Kohli (INDIA)	MS Dhoni (INDIA)	V Kohli (INDIA)
2	RT Ponting (AUS)	RT Ponting (AUS)	AD Mathews (SL)
3	MJ Clarke (AUS)	DPMD Jayawardene (SL)	GC Smith (SA)
4	KC Sangakkara (SL)	GC Smith (SA)	RT Ponting (AUS)
5	GC Smith (SA)	AB de Villiers (SA)	DL Vettori (NZ)
6	EJG Morgan (ENG)	V Kohli (INDIA)	MS Dhoni (INDIA)
7	BB McCullum (NZ)	MJ Clarke (AUS)	Shakib Al Hasan (BDESH)
8	Asghar Stanikzai (AFG)	DL Vettori (NZ)	AB de Villiers (SA)
9	AB de Villiers (SA)	Shakib Al Hasan (BDESH)	MJ Clarke (AUS)
10	MS Dhoni (INDIA)	Misbah-ul-Haq (PAK)	DPMD Jayawardene (SL)
11	DPMD Jayawardene (SL)	KS Williamson (NZ)	Mortaza (BDESH)
12	MP Vaughan (ENG)	BB McCullum (NZ)	KS Williamson (NZ)
13	Mortaza (BDESH)	KC Sangakkara (SL)	MP Vaughan (ENG)
14	DL Vettori (NZ)	EJG Morgan (ENG)	SO Tikolo (KENYA)
15	AN Cook (ENG)	AD Mathews (SL)	CH Gayle (WI)
16	Misbah-ul-Haq (PAK)	SO Tikolo (KENYA)	P Utseya (ZIM)
17	KS Williamson (NZ)	WTS Porterfield (IRE)	DJG Sammy (WI)
18	SPD Smith (AUS)	MP Vaughan (ENG)	E Chigumbura (ZIM)
19	AD Mathews (SL)	Mortaza (BDESH)	JO Holder (WI)
20	WTS Porterfield (IRE)	SPD Smith (AUS)	SPD Smith (AUS)
21	Shakib Al Hasan (BDESH)	AN Cook (ENG)	KC Sangakkara (SL)
22	AJ Strauss (ENG)	AJ Strauss (ENG)	BB McCullum (NZ)
23	Habibul Bashar (BDESH)	P Utseya (ZIM)	EJG Morgan (ENG)
24	DJG Sammy (WI)	DJG Sammy (WI)	Misbah-ul-Haq (PAK)
25	SO Tikolo (KENYA)	CH Gayle (WI)	WTS Porterfield (IRE)
26	CH Gayle (WI)	E Chigumbura (ZIM)	AN Cook (ENG)
27	JO Holder (WI)	JO Holder (WI)	AJ Strauss (ENG)
28	P Utseya (ZIM)	Habibul Bashar (BDESH)	Asghar Stanikzai (AFG)
29	E Chigumbura (ZIM)	Asghar Stanikzai (AFG)	Habibul Bashar (BDESH)

It has always been said that M S Dhoni of India is a very great captain but if only winning percentage is considered, he is ranked 10th. He is ranked 6th using weighted average method while is ranked 1st using Principal Component Analysis. Ricky Ponting of Australia is ranked 2nd using winning percentage and Principal Component Analysis while is ranked 4th using weighted average method.

Conclusions

Measuring athletic performance is an exciting task in any sport. It is especially important in competitive sports like cricket which is impacted by player actions which, by their nature, usually involve organizations spending large monetary sums with the hope of future, competitive advantages. The role of the captain in cricket directly impacts the result of the match and eventually the profit of the franchise. Unfortunately, there is no measure to quantify the captaincy.

Although expert opinion can be quite valuable; it is also very subjective. Here, we have demonstrated a simple method using principal component analysis that can be directly applied to correlated, multivariate data. Using ODI data, we have shown how to rank captains based on their individual contributions to their teams and also the team's performance under their captaincy.

References

1. Damodaran U. Stochastic dominance and analysis of ODI batting performance: The Indian cricket team, 1989-2005. Journal of Sports Science and Medicine. 2006; 5:503-508.
2. Kimber AC, Hans ford AR. A statistical analysis of batting in cricket. Journal of Royal Statistical Society. 1993; 156:443-455.
3. Lemmer H. The combined bowling rate as a measure of bowling performance in cricket. South African Journal for Research in Sport, Physical Education and Recreation. 2002; 24(2):37-44.
4. Lemmer H. A measure for the batting performance of cricket players. South African Journal for Research in Sport, Physical Education and Recreation. 2004; 26(1):55-64.
5. Lemmer H. A method for the comparison of the bowling performances of bowlers in a match or series of matches. South African Journal for Research in Sport, Physical Education and Recreation. 2005; 27(1):91-103.
6. Lemmer H. A measure of the current bowling performance in cricket. South African Journal for Research in Sport, Physical Education and Recreation. 2006; 28(2):91-103.
7. Lemmer H. An analysis of players' performances in the first cricket Twenty 20 world cup series. South African Journal for Research in Sport. 2008; 30:71-77.
8. Lenten LJA, Geerling W, Kónya L. A Hedonic Model of Player Wage Determination from the Indian Premier League Auction Seminar Paper, Department of Economics and Finance, La Trobe University, Australia, 2009.
9. Manage Ananda BW, Scariano Stephen M. An Introductory Application of Principal Components to Cricket Data, Journal of Statistics Education. 2013; 21(3):1-22
10. Mukherjee S. Identifying the greatest team and captain - A complex network approach to cricket matches. Physica A: Statistical Mechanics and its Applications. 2012; 391(23).
11. Naik DN, Khattree R. Revisiting Olympic Track Records: Some Practical Considerations in the Principal Component Analysis. The American Statistician. 1996;

50(2):140-144.

12. Radicchi F. Who Is the Best Player Ever? A Complex Network Analysis of the History of Professional Tennis. PLoS ONE. 2011; 6(2):e17249

13. Shah Parag. New performance measure in Cricket, IOSR Journal of Sports and Physical Education. 2017; 4(3):28-30.

14. Tan A, Ramachandran R. Ranking the greatest all-rounder's in test cricket, 2010.
Available at www.cricketsociety.com/ranking_the_greatest_all-ro.pdf, Retrieved on October 17.

15. Van Staden PJ. Comparison of Bowlers, Batsmen and All-rounders in Cricket Using Graphical Display, Technical Report 08/01, Department of Statistics, University of Pretoria, South Africa, 2008.

16. Van Staden PJ. Comparison of Cricketers' Bowling and Batting Performances Using Graphical Displays, Current Science. 2009; 96(6):764-766.

17. Wood GH. Cricket scores and geometric progression. Journal of Royal Statistical Society. 1945; 108:12-22.