Techno-Science

Scientific Journal of Mehmet Akif Ersoy University

www.dergipark.gov.tr/sjmakeu

Original Research Article

hosted by Turkish JournalPark ACADEMIC

Performance Analysis of Pakistan Super League Players Using Principle Component Analysis Approach

Ishrat Riaz[1*], Noureen Mushtaq[1], Mir Muhammad Jillani[2], Uzma Navaz[1]

[1] *Department of Statistics, The Women University Multan, Pakistan*
[2] *Data Centre Multan Electric Power Company (MEPCO), Pakistan*

ARTICLE INFO

Article History
Received : 20/11/2019
Revised : 30/12/2019
Accepted : 30/12/2019
Available online : 30/12/2019

Keywords
T-20
Batting
Bowling
PCA
PSL
Ranking

ABSTRACT

Where there is sport there is statistics and cricket is no exception to this. The game of cricket has a wide wealth of complex statistical data associated with the game. This study provides an outstanding application of Principle Component Analysis in evaluating the performance analysis of Cricket data. This study probes the systematic covariation among various dimensions pertaining to batting and bowling capabilities of Players of Pakistan Super League PSL T-20 (2016-2017) using the advanced statistical technique Principle Component Analysis. In the present study PCA is used to rank the batsmen and bowlers of PSL based on their contributions to their teams during these competitive seasons. The findings of this study showed the best top ten ranked batsmen and bowlers who performed well during the series also we can concluded that batting capability dominates over bowling capability. This conclusion coincides with the general opinion of several cricketing enthusiasts and experts. This research is a first study in Pakistan that highlights the features of PSL.

1. INTRODUCTION

Cricket or the gentleman's game is a very old, widespread and uncomplicated pastime game. In the late 16th century, the sport of cricket has originated in the southeast England. It became the country's National sport in the 18th century and has developed globally in the 19th and 20th centuries and yet the most popular game of the today's world. It is a game of uncertainty. One cannot predict outcome of the game up to the last moment of the game though the possible results are known to all, therefore, an appropriate probability model can be applied to predict the result Shah, Hazarika, and Hazarika (2017) [1].

Twenty 20 cricket is a form of limited overs cricket which has gained popularity worldwide. Twenty20 cricket was showcased in 2003 and involved matches between English and Welsh domestic sides [2]. The rationale behind the introduction of T20 was to provide an exciting version of cricket with matches concluding in three hours duration or less. There are now various professional T20 competitions where our concern is with Pakistan Super League (PSL).

The Pakistan Super League (PSL) is a professional Twenty20 cricket league in Pakistan composed of 5 member clubs. Headquartered in Lahore, the PSL is considered to become the premier professional cricket league in Pakistan. The league is composed of 5 competing teams Peshawar Zalmi, Quetta Gladiators, Islamabad United, Lahore Qalandars and Karachi Kings. The PSL regular season runs through the month of February, with each team playing 8 games; the top 4 teams with the best record will qualify for the post season (or playoffs), culminating in the championship game, the PSL Cup. Instead of operating as an association of independently owned teams, the PSL is a single entity in which each team is owned and controlled by the league's investors. In September 2015, the Pakistan Cricket Board officially announced the launch of the Pakistan Super League. Former captains Wasim Akram and Rameez Raja signed up to become brand ambassadors of the PSL for the next three years. After several years of planning, the league will officially begin on February 4, 2016, in the United Arab Emirates. The PSL immediately took the place as one of the leagues that contested for the annual T20 cup. The

* Corresponding Author: Ishrat.riaz@wum.edu.pk

To cite this article: Riaz, I., Mushtaq, N., Jillani, M.M., Navaz, U., (2019). Performance Analysis of Pakistan Super League Players Using Principle Component Analysis Approach. Techno-Science, vol. 2, no. 4, p. 127-135.

e-ISSN 2651-3722

PSL will begin with five teams (all based in Pakistan) and, through a series of expansions over the next several seasons, will expand to a 10 team league with two conferences.

The league draws many highly skilled players from all over the world and currently has players from approximately 11 different countries. Pakistanis have historically constituted the majority of the players in the Haier T20 League in the PSL however there is a higher percentage of Asian and European players. The league uses the draft system for player recruitment, similar to that of the National Football League in the United States, as opposed to the "auction" used in other T20 leagues. The PSL's official logo and anthem of the first season were launched on 20 September 2015 in a launching ceremony in Lahore. Basically in this study we quantify the athletic performance through PCA.

- This research aims to use Principle Component Analysis to rank the batsmen and bowlers of PSL season 1 and 2 (2016- 2017), So that we are able to identify the top batsmen and bowlers of PSL competition
- To evaluate the batting capabilities over bowling capabilities in PSL.

2. THE DATA AND METHODOLOGY

This study explored secondary data and the source of the data is from the Cricinfo website, www.espncricinfo.com from which the required data has been extracted for PSL Season I and II (2016-2017), which are freely available in the above mention website. In this study a set of 50 batsmen (who played at least 4 innings in the PSL-I-II) and 48 bowler's (who bowled at least 4 overs in the PSL-I-II) has been considered respectively.

Principle Component Analysis (PCA) is a non parametric technique used to collapse a set of correlated variables into fewer uncorrelated variables as linear combinations of the original variables [3]. Here, the purpose of using PCA is to rank the batsmen and bowlers of Pakistan Super League (PSL) 2016-2017. The study of batting and bowling performances involve eight important batting Parameters and four bowling parameters. Descriptions of these different parameters of batting and bowling performances are given in the Table 1 and 2 below:

Table 1. *Batting Parameters*

Batting Parameters	Description
Runs	The total runs scored by a player in the PSL. Higher values indicate stronger performance.
Highest individual score (HS)	The highest score ever made by the batsman in PSL.
Batting average (Ave)	The total number of runs a batsman has scored divided by the total number of times he has been called out in PSL.
Batting strike rate (SR)	The Batting Strike Rate is defined as the number of runs scored per 100 balls faced by a batsman in PSL.
Ball faced (BF)	Total number of balls received, including no-balls but not including wide in PSL.
Fifties (50)	The number of innings in which the batsman scored fifty to ninety-nine runs in PSL.
4's	The total number of boundaries (fours = four runs) made in the PSL by a batsman.
6's	The total number of sixes (= six runs) made in the PSL by a batsman.

Table 2. *Bowling Parameters*

Bowling Parameters	Description
Wickets	The number of wickets taken by a bowler in PSL.
Bowler's economy rate	The average number of runs conceded per wicket in PSL.(Eco= Runs/Wkts)
Bowling average	The average number of balls bowled per wicket taken in PSL. (SR = Balls/Wkts)
Bowling strike rate	The average number of runs conceded per over in PSL. (Econ = Runs/(overs bowled)

Cricket analytics was done using MINITAB 18 version. Team and player ratings are typically a function of some quantitative estimate of performance in the sporting arena. The performance analysis in cricket i.e. the batting and bowling performances of the players of Pakistan Super League (PSL, 2016-2017) has been evaluated using Principle Component Analysis (PCA).

3. RESULT AND DISCUSSION

Results of Batting Performance

For evaluating the batting performance of the players in PSL, the eight batting parameters: runs, highest individual score (HS), average batting performance (Ave), strike rate (SR), numbers of fours (4's), ball faced (BF), number of fifties (50) and number of sixes (6's) are used.

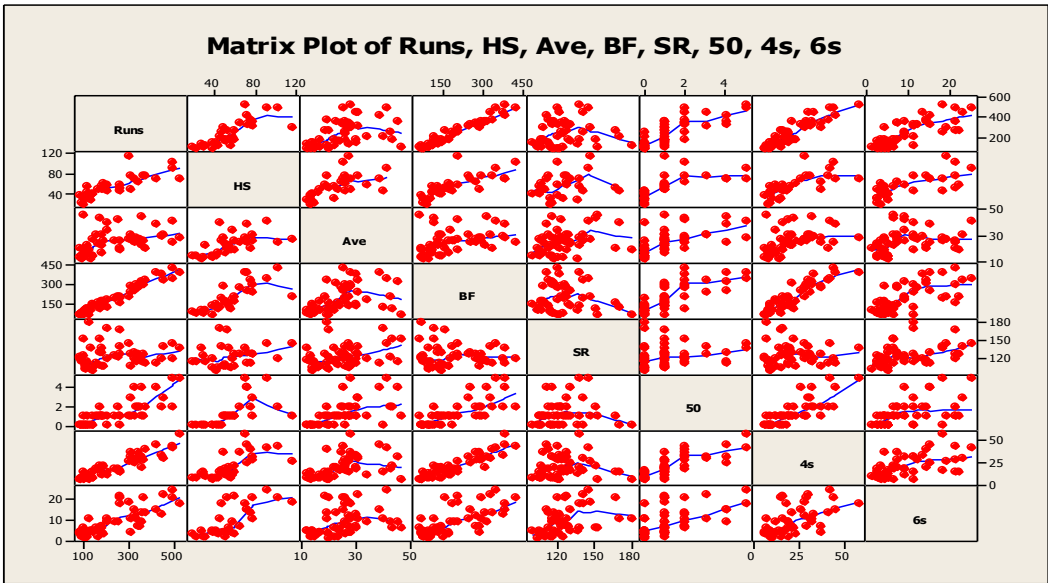

Fig. 1. *Matrix plot of eight batting parameters.*

Figure 1 illustrates a matrix plot; it shows that there are significant correlations exist among these measures. Runs and HS, Runs and BF, Runs and 4s and Runs and 6s are significantly correlated.

In order to study batting performance, PCA technique which is capable of handling correlated data in any reasonable attempt is adopted. PCA Analysis used these Batting Measures i.e. Runs, HS, Ave, BF, SR, 50, 4s, and 6s with the condition that all the batsmen who played at least 4 matches in PSL are considered. This accounts 50 players. The number of innings played for a batsman is the number of games in which he actually bats; however, in limited-overs cricket, the game could conclude before a batsman ever gets to bat, which would not count as an innings for that particular batsmen. The sample correlation matrix examined for the correlation structure inherent in these cricket variables, because these variables are measured on very different scales, they must be standardized before PCA analysis. However, the process of finding the principal components by using the standardized variables is equivalent to finding principal components by using the correlation matrix instead of the covariance matrix.

Table 3 indicates the sample correlation matrix for the fifty batting vectors. Here for computing the Eigen value-eigenvector pairs of a sample correlation matrix we use Minitab.

Table 3. *Sample Correlation Matrix of Batting Statistics*

	Runs	HS	Ave	BF	SR	50	4s	6s
Runs	1	0.795	0.447	0.974	0.150	0.801	0.916	0.786
HS	0.795	1	0.581	0.764	0.254	0.624	0.702	0.728
Ave	0.447	0.581	1	0.400	0.336	0.594	0.443	0.325
BF	0.974	0.764	0.400	1	-0.048	0.767	0.914	0.681
SR	0.150	0.254	0.336	-0.048	1	0.124	0.064	0.467
50	0.801	0.642	0.594	0.767	0.124	1	0.812	0.552
4s	0.916	0.702	0.443	0.914	0.064	0.812	1	0.586
6s	0.786	0.728	0.325	0.681	0.467	0.552	0.586	1

Table 4 represents the ordered Eigen values and percentage of total variability attributed to each Eigen value. This revels that first PC explains the most of variation that is 68.2% against the first Eigen value. So, we take first Principle Component to rank the batsmen.

Table 4. *Ordered Eigen Values and Corresponding Percentages of Total Variability*

Eigen values	Total variability
5.4573	68.2%
0.8849	14.8%
0.5436	6.8%
0.3749	4.7%
0.2240	2.8%
0.1892	2.4%
0.0182	0.2%
0.0034	0.0%

Now, the variables Runs, HS, Ave, BF, SR, 50, 4s, and 6s are understood to have already been individually standardized. The first entry in the 2nd row of Table 4 shows 68.2% of total variability can be explained by first principle component analysis. Also its corresponding Eigen value 5.4573 is the only one that is greater than 1. [4] suggests retaining all principal components whose corresponding Eigen values exceed unity.

Table 5 shows the eigenvector coefficients for all eight principal component (PC1 – PC8), which are listed only for completeness. The Eigenvalue-eigenvector pair for the first principal component is highlighted in Table 5.

Table 5. *Eigen Values and Eigen Vectors for Sample Correlation Matrix*

Eigen values	5.4573	0.8849	0.5436	0.3749	0.2240	0.1892	0.0182	0.0034
Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Runs	0.420	0.080	0.117	0.149	0.199	0.081	-0.374	0.771
HS	0.364	0.006	0.271	-0.660	-0.279	-0.528	-0.042	-0.006
Ave	0.316	-0.195	-0.779	-0.409	0.156	0.240	0.076	0.017
BF	0.399	0.278	0.114	0.051	0.322	0.149	-0.496	-0.613
SR	0.160	-0.186	-0.037	0.345	0.032	-0.349	-0.213	-0.139
50	0.366	0.193	-0.233	0.378	-0.787	0.081	-0.014	-0.069
4s	0.388	0.242	-0.062	0.319	0.362	-0.401	0.626	-0.037
6s	0.349	-0.341	0.483	-0.088	-0.049	0.589	0.410	-0.060

Figure 2 represents the ordered Eigen values plotted sequentially; the resulting visual presentation is commonly called a Scree plot, which may be used to ascertain the appropriate number of principal components to retain in a particular application. To do this, one looks for an elbow (bend) in the plot. The number of useful principal components is then taken to be the abscissa of the point beyond which all remaining Eigen values add relatively small contributions to the total variability.

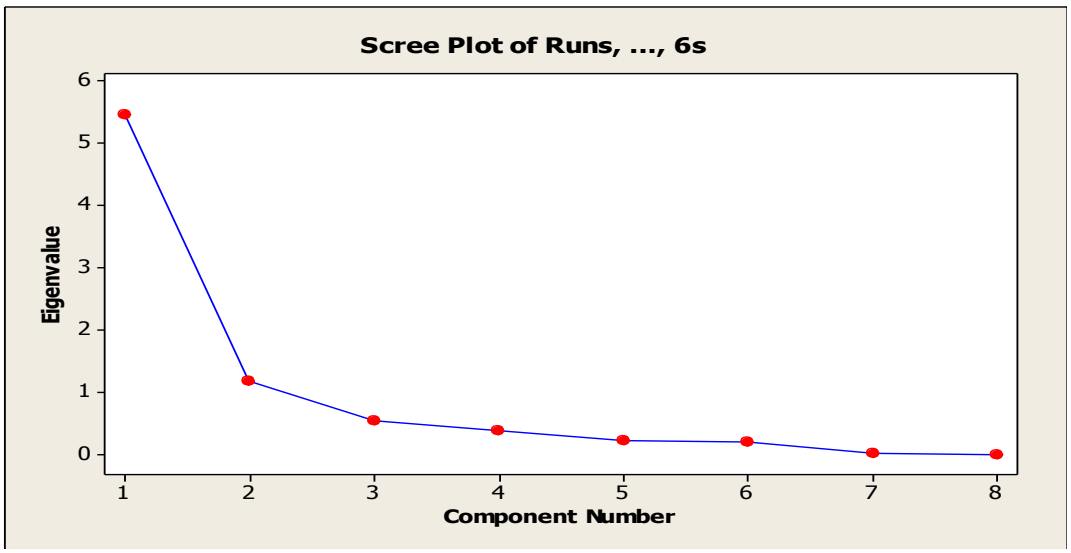

Fig. 2. *Scree Plot for Batting Performance*

Figure 2 shows the Scree plot for the Batting statistics, suggesting that since the elbow is at abscissa two, it is reasonable to use only the first principal component which explains 68.2% of the total variability. Hence as a result the first principle component analysis for batsmen is:

$$L_1 = (0.42*Runs)+(0.364*HS)+(0.316*Ave)+(0.399*BF)+(0.16*SR)+(0.366*50)+(0.388*4s)+(0.349*6s)$$

Ranking Batsmen Using First Principle Component Analysis

Here, it is to be referred the first principal component as the general-batting-performance-index, which is a type of weighted average of all eight variables used. Here, the coefficients of the first principal component are all positive, so larger values of L_1 indicate better player performance. This justifies that we should rank (largest to smallest) the players based on the first principal component. Table 6 indicates the top 10 batsmen who played at least 4 matches in PSL (2016-2017) using first principle component L_1.

Table 6. *Top ten Batsmen of PSL Ranked by First Principal Component, L_1*

Ranking	Batsmen	Matches	Inns	Runs	HS	Ave	BF	SR	50	4s	6s	L_1
1	Kamran Akmal	21	20	504	104	25.2	420	120	2	44	21	469.41
2	Ahmad Shehzad	20	19	532	71	28	385	138.18	5	58	18	464.47
3	Umar Akaml	15	15	499	93	41.58	345	144.63	5	42	25	444.22
4	DR Smith	13	13	427	73	38.81	385	110.9	4	45	15	413.69
5	RS Bopara	19	18	448	71*	29.86	372	120.43	2	33	13	409.21
6	KP Pietersen	18	17	456	88*	30.4	324	140.74	3	30	23	405.72
7	Mohammad Hafeez	20	19	373	77	19.63	333	112.01	2	41	14	363.21
8	Tamim Iqbal	11	11	361	80*	45.12	313	115.33	4	32	9	355.36
9	SR Watson	15	15	365	79	26.07	290	125.86	1	29	21	345.09
10	Shoaib Mailk	19	19	359	51*	22.43	308	116.55	4	32	9	330.98

While Table 7 shows a complete list of PSL (2016-2017) batsmen ranked using first PC. Here, the '*' indicates that a player not out in a match.

Table 7. *Complete List of PSL Batsmen Ranked by First Principal Component, L_1*

Batsmen	Ranking	L_1	Batsmen	Ranking	L_1	Batsmen	Ranking	L_1
Kamran Akmal	1	469.41	Sharjeel Khan	18	299.89	Mohammad Nawaz	35	155.79
Ahmad Shehzad	2	464.47	RR Rossouw	19	261.35	GD Elliott	36	152.24
Umar	3	444.22	CH Gayle	20	255.02	Sohaib Maqsood	37	139.72
DR Smith	4	413.69	DJG Sammy	21	244.07	BJ Hodge	38	138.21
RS Bopara	5	409.21	Shahid Afridi	22	239.62	Imad Wasim	39	136.26
KP Pietersen	6	405.72	Mohammad Rizwan	23	229.57	JM Vince	40	130.84
Mohammad Hafeez	7	363.21	Khalid Latif	24	214.69	DJ Bravo	41	126.69
Tamim Iqbal	8	355.36	LI Wright	25	204.22	SP Narine	42	124.99
SR Watson	9	345.09	Azhar Ali	26	202.92	Sohail Tanvir	43	119.87
Shoaib Malik	10	330.98	KA Pollarad	27	198.06	Shahid Yousuf	44	116.54
DJ Malan	11	325.72	Shakib-ul-Hasan	28	189.54	BB McCullum	45	110.44
Sarfraz Ahmad	12	324.18	JJ Roy	29	188.91	Anwar Ali	46	106.84
KC Sangakkara	13	321.55	SW Billings	30	186.33	AD Russell	47	103.67
Misbah-ul-Haq	14	310.15	Fakhar Zaman	31	184.74	MN Samuels	48	102.95
BJ Haddin	15	308.94	Asad Shafiq	32	179.66	Wahab Raiz	49	101.94
Babar Azam	16	302,52	LMP Simmons	33	176.74	Asif Ali	50	101.18
CS Delport	17	301.89	EJG Morgan	34	170.91			

Results of Bowling Performance

Four bowling measures such as wickets, bowler's economy rate, bowling average, and bowling strike rate has been used for evaluating the bowling performance of the players in PSL. The correlation structure of the four "bowling variables"

which we use is shown in the matrix plot of Figure 3. The figure shows that Bowling Average and the Strike Rate are highly positively correlated. All the other variables are somewhat negatively correlated with the number of wickets. However, each one of these variables measures a different quality of a bowler, even though they are correlated.

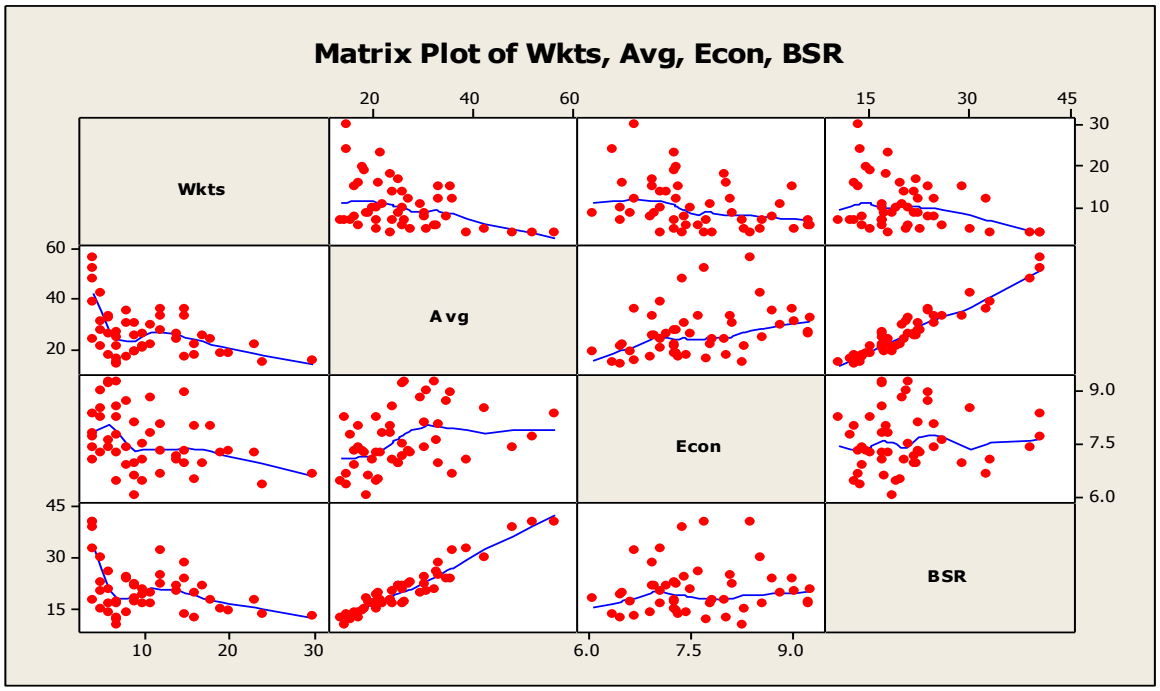

Fig. 3. *Matrix Plot for Bowling Variables*

Again in order to study bowling performance, PCA technique which is capable of handling correlated data in any reasonable attempt is adopted. The variables most commonly used for measuring the quality of bowlers are Wickets (Wkts), Bowling Average (Ave), Bowling Strike Rate (BSR), and Economy Rate (Econ). Bowlers who have bowled at least 4 overs are considered, which accounts for 48 total bowlers.

We use the sample correlation matrix in an attempt to understand the correlation structure of the bowling vectors. Table 8 shows the sample correlation matrix for the 48 bowling observations.

Table 8. *Sample Correlation Matrix for Bowling Performance*

	Wkts	Ave	Econ	SR
Wkts	1	-0.425	-0.366	-0.353
Ave	-0.425	1	0.357	0.955
Econ	-0.366	0.357	1	0.075
SR	-0.353	0.955	0.075	1

The Given Table 9 provides the ordered Eigen values and percentage of total variability against each Eigen values. While Table 10 gives the coefficients for all four principal component (PC1 – PC4). The Eigen value Eigen vector pair for the first principal component is highlighted in Table 10.

Table 9. *Ordered Eigen Values and Corresponding Percentages of Total Variability*

Eigen Values	2.3495	0.9586	0.6089	0.0030
Total Variability	58.7%	26.0%	15.2%	0.10%

Table 10. *Eigen Values and Eigen Vectors for Sample Correlation Matrix*

Eigen Values	2.3495	0.9586	0.6089	0.0030
Variable	PC1	PC2	PC3	PC4
Wkts	-0.431	0.397	-0.810	0.010
Ave	0.624	0.226	-0.230	-0.712
Econ	0.314	-0.756	-0.535	0.208
BSR	0.572	0.469	-0.066	0.670

Accordingly, the first principal component for bowlers is

$$L_1 = -0.431*Wkts + 0.624*Ave + 0.314*Econ + 0.572*BSR$$

The first entry in Table 9 shows that 58.7% of the Total Variability can be explained against first Eigen value by this first principal component alone. The associated Scree plot is shown in Figure 4, which indicates an elbow at abscissa 2, implying that the first principal component, is sufficient for ranking bowlers. Also its corresponding Eigen value 2.3495is the only one which is greater than 1.

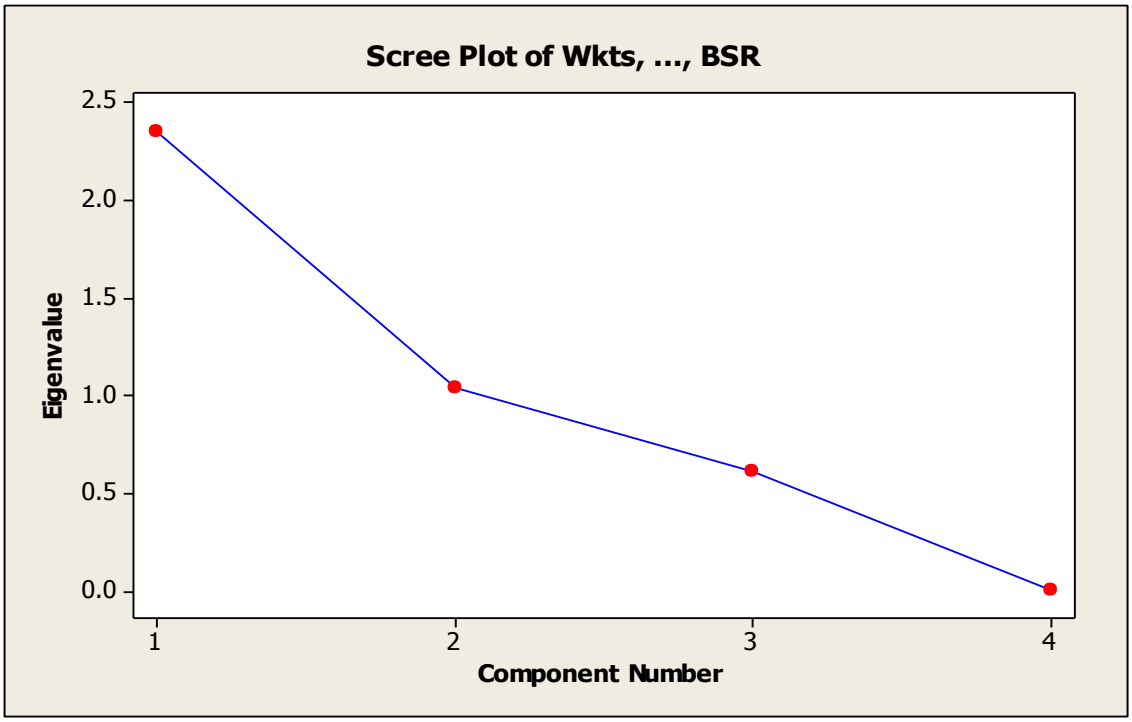

Fig. 4. *Scree Plot for Bowling Performance*

Ranking Bowlers Using First Principle Component Analysis

Since, L_1 = -0.431*Wkts + 0.624*Ave + 0.314*Econ+0.572*BSR is a type of weighted average of these four variables, it is reasonable to use the first principle component for ranking. L_1 can reasonably be described as the general-bowling-performance-index, where smaller (negative) values indicate stronger bowler performance. Unlike the analysis for batsmen, where the coefficients in L_1 of that discussion are all positive, the coefficient of Wkts here is negative while the all other remaining are positive. This makes sense since better bowler performance is naturally associated with higher numbers of wickets taken from batsmen. On the other hand, lower values for the other three variables: Bowling Average (Ave), Economy Rate (Econ) and Bowling Strike rate (BSR) indicate better bowler performance. Table 11 lists the top ten bowlers using the first principal component L_1. Table 12 shows a complete list of PSL bowlers (2016-2017).

Table 11. *Top Ten Bowlers of PSL Ranked by First Principal Component, L_1*

Ranking	Bowler	Matches	Wickets	Ave	Econ	SR	L_1
1	Wahab Raiz	19	30	14.93	6.65	13.4	6.14
2	Mohammad Sammi	16	24	14.75	6.35	13.9	8.80
3	Sohail Khan	14	20	18	7.29	14.8	13.37
4	AD Russell	10	16	17.25	8.03	12.8	13.71
5	GD Elliot	13	15	16.6	7.32	13.6	13.97
6	Mohammad Asghar	16	19	18.31	7.25	15.1	14.15
7	Zafar Gohar	6	7	14.57	8.27	10.5	14.68
8	Mohammad Hafeez	20	7	13.85	6.46	12.8	14.98
9	Mohammad Nawaz	20	23	21.65	7.25	17.9	16.11
10	Mohmuduallah	6	7	15.85	7.27	12.2	16.28

Table 12. *Complete List of PSL Bowlers Ranked by First Principal Component, L_1*

Bowlers	L_1	Rank	Bowlers	L_1	Rank	Bowlers	L_1	Rank
Wahab Raiz	6.14	1	Yasir Shah	20.37	17	Aamer Yamin	30.39	33
Mohammad Sammi	8.80	2	RS Bopara	21.10	18	Saeed Ajmal	30.40	34
Sohail Khan	13.37	3	SP Narine	21.62	19	NLTC Perera	31.80	35
AD Russell	13.71	4	TS Mills	21.96	20	Shakib-ul- Hasan	32.03	36
GD Elliot	13.97	5	CS Delport	22.24	21	Sohail Tanvir	32.40	37
Mohammad Asghar	14.15	6	Hasan Ali	22.72	22	Anwar Ali	32.41	38
Zafar Gohar	14.68	7	Mohammad Amir	23.29	23	DJG Sammy	32.62	39
Mohammad Hafeez	14.98	8	Ehsan Adil	24.55	24	Zulfiqar Babar	32.94	40
Mohammad Nawaz	16.11	9	Mohammad Irfan	24.95	25	Junaid Khan	34.77	41
Mohmuduallah	16.28	10	BAW Mendis	25.69	26	Usman Khan	35.27	42
CJ Jordan	17.05	11	Mohammad Irfan (Jnr)	26.15	27	Shahid Afridi	38.02	43
Aizaz Cheema	18.57	12	KK Cooper	26.36	28	Shoaib Malik	43.54	44
Usama Mir	19.49	13	SW Tait	26.40	29	Umar Gul	44.26	45
Rumman Raees	19.75	14	Hassan Khan	26.53	30	S Badree	52.85	46
Imran Khalid	19.99	15	Imad Wasim	27.04	31	Mohammad Nabi	56.31	47
Shadab Khan	20.02	16	SR Watson	28.01	32	DJ Bravo	59.33	48

4. CONCLUSIONS

The need for insertion of advanced research methods has increased in the past few years due to the massive investments and profit-making importance of sports. Many professional sports have enjoyed the attention of deep-rooted research methods; cricket still continues to be a sport that has received satisfactory attention.

This research is an imperative footstep in the precise direction. The research has used principal component analysis for the first time in cricket game research problem of ranking the PSL players on the basis of their batting and bowling statistics. Principle component analysis used to rank the batsmen and bowlers of PSL (2016-2018). The principle component analysis investigates the interrelationship among various dimensions of PSL. Data of 50 batsmen and 48 bowlers with a variety of dimensions of batting and bowling were used. The eight dimensions have been grouped for batting and four dimensions for bowling.

To the best of researcher knowledge this research is a first study in Pakistan that highlights the features of PSL and addresses the problem of ranking of players by evaluating bating and bowling performance. Ranking players in competitive, team sports is a in competitive, team sports is a challenging and elusive problem that sometimes can be accomplished using multivariate statistical techniques.

To study batting performance eight battings measures are used and in a uni-variate sense, high values for each of these variables indicates better batting performance and each one measures a different quality of a batsman. But, this study primary concern is their joint contribution to batting performance in a multivariate sense. Similarly the case is with

bowling performance so constructing an overall measure of batting and bowling performance by collapsing these correlated variables is a key goal of this paper and this measure is principal component analysis.

Result of principal component analysis also indicates that the variance explained by batting is much higher than bowling which shows apparent supremacy of batting capability over bowling capability. This was an essential feature of cricket frequently used by team selectors. The result of research provides higher priority to batting capability over bowling capability.

REFERENCES

[1]. Shah, S., Hazarika, P. J., & Hazarika, J. (2017). A study on performance of cricket players using factor analysis approach. *International Journal of Advanced Research in Computer Science*, 8(3).

[2]. Perera, H., Davis, J., & Swartz, T. B. (2016). Optimal lineups in Twenty20 cricket. *Journal of Statistical Computation and Simulation*, 86(14), 2888-2900.

[3]. Richardson, M. (2009). Principal component analysis. URL: http://people. maths. ox. ac. uk/richardsonm/SignalProcPCA. pdf (last access: 3.5. 2013). Aleš Hladnik Dr., Ass. Prof., Chair of Information and Graphic Arts Technology, Faculty of Natural Sciences and Engineering, University of Ljubljana, Slovenia ales. hladnik@ ntf. uni-lj. si, 6, 16.

[4]. Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, 20(1), 141-151.