Statistical Analysis of Cricket Leagues Using Principal Component Analysis

Sheharyar Khan [1], Muhammad Ishtiaq [2], Kainat Bibi [3],Rashid Amin[*4] Adeel Ahmed[5], Iqra Chaudhary[6]

ABSTRACT-- *Any sport has statistics, and cricket is the one, where statistics are extremely important because players are ranked using these data. Individual runs, wickets, and highest scores, among other things, are included in these statistics. Players are chosen for tournaments all over the world based on statistics. By analysing cricket statistics and figures, this study employs Principal Component Analysis. Using the approach called Principal Component Analysis, this study examines the precise co-variation among several measurements linked to the batting and bowling talents of players in the Pakistan Super League PSL T-20 (2016-2019) and the Indian Premier League IPL T-20 (2016-2019). PCA is applied in this study to rank the PSL batsmen and bowlers based on their contributions to their clubs during these competitive seasons. The results of this research revealed the top ten ranked batters and bowlers who excelled during the series. Principal Component Analysis is widely used in applied multivariate data analysis. In the current investigation, PCA was utilized to rank the top ten best-performing batsmen and bowlers of the PSL and IPL. Principal Component Analysis is a dimension reduction technique that is used to reduce dataset dimensions into smaller variables. Here, principal component analysis is successfully used to rank the cricket batsmen and bowlers.*

Mathematics Subject Classification (2010) 62H25, 62H86, 62J10

Keywords PSL, IPL, PCA, Sports Analysis

I. INTRODUCTION

Cricket is a bat-and-ball sport played between two teams in which each team includes eleven players. A 20-meter-long pitch with a wicket at each end is located in the centre. A bail is attached to each wicket. Batsmen score runs by hitting the bowler's ball, and bowlers attempt to take wickets by hitting the ball to the wickets. Other methods of taking wickets include being caught out by any fielder on the ground when batsmen smash a ball, getting stumped out by the keeper behind the batsman, or hitting a wicket with the bat when batsmen strike their bat to the wicket.

*Corresponding Author
Email addresses:
sheharyar.khan@uettaxila.edu.pk,ishtiaqmohammed31 @gmail.com, kiynat001@gmail.com,
*rashid.amin@uettaxila.edu.pk,
Received: DD.MM.YYYY; Accepted: DD.MM.YY

Similarly, batsmen scores by hitting boundaries. When the bowling side dismisses ten players, the innings of one team that's batting ends, and the bowling team sends their batsman to bat. The first batting team's target is now being chased by the second batting team. Based on the chase or dismissed by the bowling team, win and loss are decided. The game is supervised by two umpires, aided by a third umpire, usually called the TV Umpire, and a match referee in international matches. They communicate with two off-field scorers who record the match's statistical information. This game is currently in three formats.

- Test-In with 5 days for play and no limits of overs.
- Twenty-Twenty in which each team plays 20 overs.
- ODI- One Day International in which 50 cricket players between two teams.

Twenty-two international games have gained popularity in the previous eight to ten years, and viewers are particularly engaged in these limited-overs games, which are played in under two hours. In 2003, Twenty 20 cricket was introduced, with matches between English and Welsh domestic teams"[1]. The T20 cricket format was cost-effective, and fans like the shorter version of the game. Everyone knows about cricket in this day and age, but for those who don't, read [2]. Many extremely competent players from all around the world compete in these leagues. The players in the IPL and PSL come from about 11 different countries, largely from Asia and Europe, according to our research. Based on this data, we quantified the performance through PCA. This research aims to find the rankings of PSL and IPL league bowlers and batsmen using PCA for the last four seasons (2016-2019). From this, we can identify the top batsmen and bowlers in the PSL and IPL competitions to evaluate the batting capabilities over bowling capabilities in the PSL and IPL. *i*

II. LITERATURE REVIEW

Bowling and batting performance comparison is conducted in a graphical methodology by [3]. Furthermore, this investigation can be used to differentiate dissimilar groups of players, for instance, aggressive batsmen, bowling all-rounders, spinners, and other groups. But graphical approaches are not accurate and there is a gap in introducing other techniques for statistical analysis. For selection and comparison[4] suggest a numerical approach through this approach to picking batsmen in the team. [5] Presents a mockup approach for the order of batsmen and batting in one-day cricket. This work was furthermore loosened up by [6]. He created a model which can simulate one-day cricket. Some advanced logical techniques, such as entire number programming and data envelopment assessment, have been applied in the composition of many components of cricket, such as player selection for the team (see Sharp et al., 2011, Lemmer [7,8]. Factor analysis based solely on global T20 data. On IPL 9, a study was conducted on the performance of cricket players using a factor analysis approach [9]. Using Principal Component Analysis, a captain's position is determined based on a few characteristics. [10] used a factor analysis of only IPL players. In one-day international cricket for Bangladesh, [11] used progressive head section-based calculation for optimal lineup and batting need assurance. Following an examination of composition specialists' work on the IPL and World T20, it was discovered that there is a gap in the market, which is filled by providing Asian T20 affiliation Analysis pushed by country boards. For example, when extending a global portfolio [12]. Valadkhani et al. (2008) adopted a factor assessment approach. According to the author's study, no one has made arrangements with Asian leagues like as the PSL and IPL in the last four seasons, and the analysis was likely done on the PSL. The restricted composition of several aspects of cricket reveals the pressing need to improve it. Statistical Method was used to rank captains based on numerous factors. The weighted average approach was also used to rank captains based on the z score of the team's performance, as well as the captain's performance as a batsman and bowler[13]. During the match, the multiple linear regression model is also used to predict the match outcome[14]. A comparative study of machine-learning methods was applied to predict cricket match outcomes using the opinions of crowds on social networks[15].

The rest of the paper is organized as follows Section III describes the Empirical Statistical Methodology employed in this paper. Section IV Steps Involved in PCA Section V Describes Data and Statistical Analysis and Section VI Introduction to IPL, Section VII Introduction to PSL, Section VIII with Concluding remarks

III. EMPIRICAL STATISTICAL METHODOLOGY

PCA is the multivariate data analysis technique devised by Pearson [16] and [17]. This technique uses mathematical principles to convert correlated variables into a smaller number of variables named Principal components. The PCA technique involves the extraction of statics from a dataset and keeping the important statistical information. PCA performs compression, simplifies the data dimensions and, reduces the dimensions of data. PCA is significant because if we perform the elimination of variables, it may cause information loss, but PCA performs extraction which can retain important information and without much loss of information, our dataset represents a more clear view of information. PCA is a technique used for correlated data and it reduces the variables [18,19,20]. Suppose there is an arbitrary vector ($\alpha = \alpha_1, \alpha_2, \alpha_3 ... \alpha_p$ consisting of p random variables, having covariance matrix and eigenvalue Eigenvector pairs $(\mu_1, e_1), (\mu_2, e_2), (\mu_3, e_3)$ (μ_p, e_p) where μ_1 $\mu_2 ...... \mu_1$ the i^{th} principal component, say Li, is defined as $Li = e^t\alpha = e_{i1}\alpha_1 + e_{i2}\alpha_2 + ... + e_{ip}\alpha_p$ $for\ i = 1, 2\ p$ where e is the component of Eigenvectore^t . Furthermore, $Z_i = b^t\alpha = b_{i1}\alpha_1 + b_{i2}\alpha_2 + ... + b_{ip}\alpha_p$ is a linear combination of variables for $Variable\ (L_1) = \mu_1\ Var\ (Z_i)\ \ Cov\ (L_i, L_j) = 0$, for $i \neq j$ then above equaitons provides that Li is the aggregate of the vital signals contained in the actual variables α. Notice the totalvariability (V_T)

$$V_T = Var(\alpha_1) + Var(\alpha_2) + \ldots + Var(\alpha_p) \quad (3.1)$$

$$V_T = Var(L_1) + Var(L_2) + \ldots + Var(L_p) \quad (3.2)$$

$$V_T = \mu_1 + \mu_2 + \ldots + \mu_p \quad (3.3)$$

For the j[th] principal component, the total percentage of variance is $\mu_j \times \sum_{i=1}^{p} \mu_i$ if a substantial percentage of the total variance is captured by the first few Principal Component Analysis. Cricket Data needs fine-tuning. If we do so, the results are accurate and does not affects the goal [21].

IV. STEPS INVOLVE IN PRINCIPAL COMPONENT ANALYSIS

a. Standardization.
b. Co-variance matrix computation.
c. Compute the eigenvectors and eigenvalues of the covariance matrix to identify the principal component.
d. Feature Vector.
e. Replicate the data along with the principal component axes.

V. DATA & STATISTICAL ANALYSIS

We have collected data from ESPN web. For analyzing the batting of players, we collected the data of those players who played at least four innings. For analyzing the bowling, we have taken the data of those bowlers who bowled at least four overs. There are 157 batsman and 135 bowlers in IPL 2016 to 2019. In adition, we have 62 bowlers and 110 batsman for PSL 2016 to 2019.

5.1. Batting Statistics

In our experimental study, we considered the measures such as highest score, total runs, average ball faced, number of the 50s, strike rate, number of 4's and number of 6's. Table 1 describe the different statistics used in our experiments for batting.

Table 1. Batting Statistics

SR No.	Bowling Statistics	Description
1	Wickets	Total Number of Wickets in Any league
2	Bowling Average	TR/W, where TR is the total runs conceded by a bowler and w is thetotal number of wickets taken by abowler
3	Bowling Economy	TR/O, where TR is the total number of runs conceded by a bowler and O is the total number of overs bowled by a bowler
4	Bowler Strike Rate	TB/W, where TB is the total number of balls bowled by a bowler and w is the total number of wickets taken by a bowler
5	Maiden Overs	Number of Over's which has no run score

Table 2. Bowling Statistics

SR No.	**Batting Statistics**	**Description**
1	Inning	Total Number of Inning a player played in all season not less than 5.
2	Total Runs (TR)	Sum of all runs in every match
3	Highest Score (HS)	The highest score by a batsman inone match during a tournament
4	Batting Average	The ratio of R/m R denotes Number Runs Score and M denotes
5	Ball Faced (BF)	Total Number of Ball Faced in allseasons and any League
6	Strike Rate	Strike Rate is TR/BF
7	No. of the 50's	Total No. of the 50's
8	No. of 4's	Total No. of 4's
9	No. of 6's	Total number of 6's

5.2. Bowling Statistics

In our experimental study, we considered the measures for bowling such as economy rate, wickets, bowling average, strike rate. Table 2 describe the different statistics used for bowling in our experiments.

VI. Introduction to IPL (Season 9-12)

Indian Premier League was first introduced by BCCI. This league is played in April to May and is most of the attended league in world. There are nine to ten teams are participating in this league in this tournament every year.

6.1. Empirical Results of IPL

6.1.1. Results of Batting Performance.

Figure 1 shows that there is a chance, due to the short format of the game, he did not bat. Similarly, a batsman who never bats is not included in our dataset. And we cannot consider it an inning. All the cricket variables are first normalized with the above rule. If we cannot do this standardization, the results are not accurate. Then we apply principal component analysis.Table 3 report the correlation matrix for 157 bats men. Table 4 shows the PC that depicts the variation of about 80% for eigenvalue.

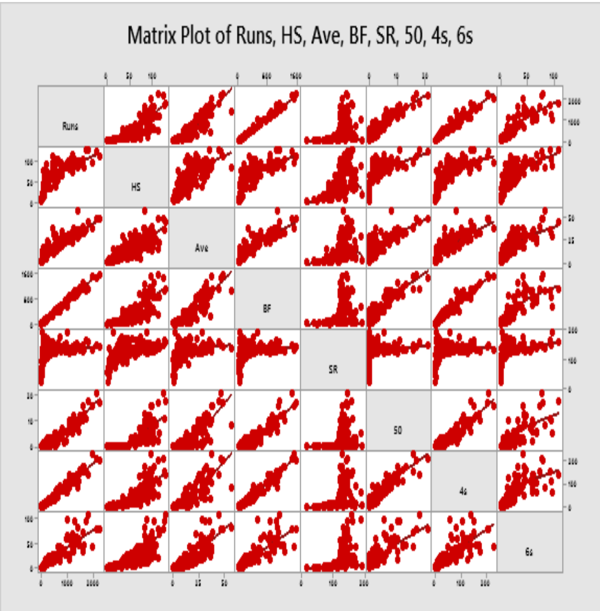

Fig. 1. Matrix plot of eight batting parameters

Table 3. Sample Correlation Matrix of Batting Statistics

VAR	RUNS	HS	AVE	BF	SR	50	4s	6s
Runs	1	0.835	0.895	0.993	0.368	0.94	0.976	0.889
HS	0.835	1	0.806	0.822	0.541	0.759	0.816	0.773
Ave	0.895	0.806	1	0.887	0.457	0.81	0.835	0.827
BF	0.993	0.822	0.887	1	0.321	0.93	0.977	0.845
SR	0.368	0.541	0.457	0.321	1	0.296	0.335	0.454
50	0.94	0.759	0.81	0.93	0.296	1	0.936	0.815
4s	0.976	0.816	0.835	0.977	0.335	0.936	1	0.802
6s	0.889	0.773	0.827	0.845	0.454	0.815	0.802	1

Table 4 shows 80.01 percent and its corresponding Eigenvalue is 6.4014. That is higher than 1 [22] recommends that retain the principal component whose corresponding Eigenvalue is greater than 1.

Table 5 below shows the Eigenvector factors for all eight principal components (PC1–PC8). The Eigenvalue Eigenvector pair for the first principal component is shown in Table 5.

Figure 2 batting statistics are shown in the scree plot. In the score plot, it is clearly shown at elbow. The point is at 2, which shows that it is used as the first principal component and it explains 80 percent of total variability. Hence, the result for batsmen is shown below:

$$L1 = (0.39 * Runs) + (0.352 * HS) + (0.365 * Ave) + (0.384 * BF) + (0.189 * SR) + (0.368 * 50) + (0.378 * 4s) + (0.358 * 6s)$$

Table 4. Eigen Values and Corresponding and total Variability

Eigen Values	**Total Variability**
6.4014	80.0175
0.9039	11.29875
0.2316	2.895
0.2026	2.5325
0.1688	2.11
0.0758	0.9475
0.0145	0.18125
0.0014	0.0175

Table 5. Eigen Values and Eigen Vectors for Sample Correlation Matrix BattingPerformance IPL

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Runs	0.390	-0.138	-0.027	0.103	0.014	0.267	-0.230	-0.833
HS	0.352	0.197	0.544	-0.617	-0.373	-0.147	-0.020	-0.010
Ave	0.365	0.044	-0.227	-0.395	0.764	-0.211	0.167	0.021
BF	0.384	-0.189	0.089	0.089	0.122	0.363	-0.627	0.511
SR	0.189	0.906	0.020	0.363	0.074	0.039	-0.055	0.012
50	0.368	-0.219	0.058	0.445	-0.112	-0.771	-0.077	0.035
4s	0.378	-0.180	0.290	0.297	0.011	0.360	0.706	0.155
6s	0.358	0.050	-0.746	-0.162	-0.493	0.072	0.135	0.141

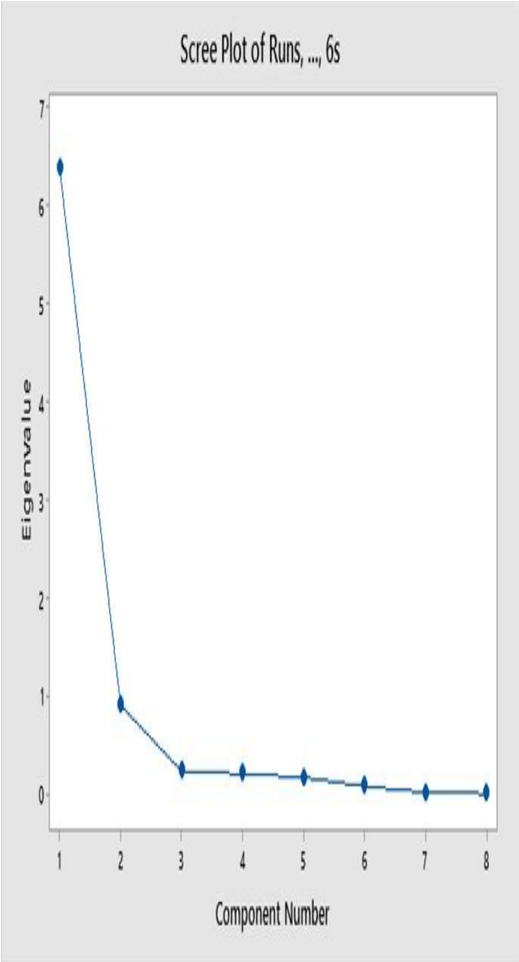

Fig. 2. Scree Plot For Batting Performance IPL

Batting Ranking of IPL

Here, it is referred to as the first principal component, which is the weighted average of all eight variables used. We have seen that the value of the first principal is positive and it is a matter of fact that the higher the value of L1 shows better performance and the lower the value of L1 shows poor performance. This gives us players ranking. Table 6 shows top ten batsmen who played at least 4 matches.

Result of Bowling Performance

Figure 3 shows the used performance variables in the matrix plot. Average and strike rates are significantly high correlation and other shows below, like maiden over, show negative impact and weaker correlation. However, each one of them shows a different type of bowler.

Table 6. Top ten Batsmen of IPL Ranked by First Principal Component, L1

Rank	Batsman	Matches	Inns	NO	Runs	HS	Ave	Bf	SR	50	4s	6s	L1
1	V Kohli	54	54	7	2275	113	48.3	1601	142.09	17	204	80	1698.30
2	DA Warma	43	43	8	2181	126	47.01	1493	146.08	21	208	78	1672.29
3	S Dhawan	63	63	9	1998	97	37.08	1552	128.73	16	227	42	1553.92
4	AB deVilhers	50	49	9	1825	129	128.73	1132	161.21	18	139	109	1336.68
5	SK Raina	61	61	8	1669	84	161.21	1269	131.52	13	172	44	1290.01
6	AM Rahana	59	57	6	1625	105	131.52	1300	125	10	173	32	1285.95
7	RP Pant	54	54	6	1736	128	125	1067	162.69	11	152	94	1270.59
8	KL Rahul	42	40	8	1649	100	162.69	1125	146.57	16	152	73	1270.599
9	RG Sharma	60	59	8	1513	94	146.57	1171	129.20	12	157	47	1241.44
10	SV Samson	55	55	6	1460	102	129.20	1082	134.93	6	110	59	1188.86

PCA out of 135 bowlers are included from these four seasons and only those are selected that bowl at least four overs. Table 7 shows the sample correlation matrix for the 135 bowling observations. In Table 7, ordered Eigenvalues and their corresponding total variability. While Table 8 shows the factors for all five principal components (PC1 – PC5). The Eigenvalue Eigenvector pair for the first principal component is highlighted in Table 8.

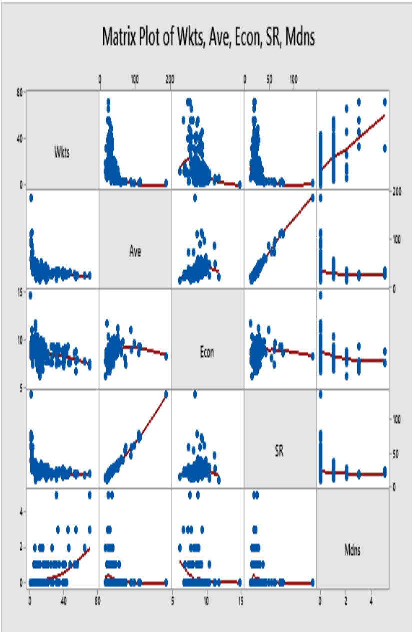

Fig.3. Matrix Plot for five Bowling parameters

Table 7. Sample Correlation Matrix for Bowling Performance

Variable	**Wkts**	**Ave**	**Econ**	**SR**	**Maiden Over**
Wkts	1	-0.425	-0.360	-0.376	0.52
Ave	-0.425	1	0.268	0.982	-0.221
Econ	-0.36	0.286	1	0.099	0.279
SR	-0.36	0.982	0.099	1	-0.181
Maiden Over	0.52	-0.221	-0.279	-0.181	1

Table 8 Eigen Values and Eigen Vectors for Sample Correlation Matrix

Variable	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**
Wkts	-0.461	0.333	0.242	0.786	0.001
Ave	0.553	0.398	0.114	0.120	0.713
Econ	0.300	-0.466	0.814	0.123	-0.124
SR	0.519	0.492	-0.032	0.107	-0.690
Maiden Over	-0.349	0.522	0.514	-0.584	-0.005

The L1 first principal component for bowlers is defined as:

$$L1 = (Wkt * (-0.461)) + (Ave * 0.553) + (Eco * 0.3) + (SR * 0.519) + (Mnd * (-0.349))$$

In Table 9 there total variability in the first entry is 51 percent and Figure 4 shown the scree plot from the scree plot it is clearly shown that the elbow at 2 gives us the first principal component and it is enough evidence for ranking bowler. Its Eigenvalue is 2.5597 and this is the only one that is higher than 1.

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

Eigen Values	**Total Variability**
2.5597	51.194
1.2541	25.082
0.7405	14.81
0.4423	8.846
0.0034	0.068

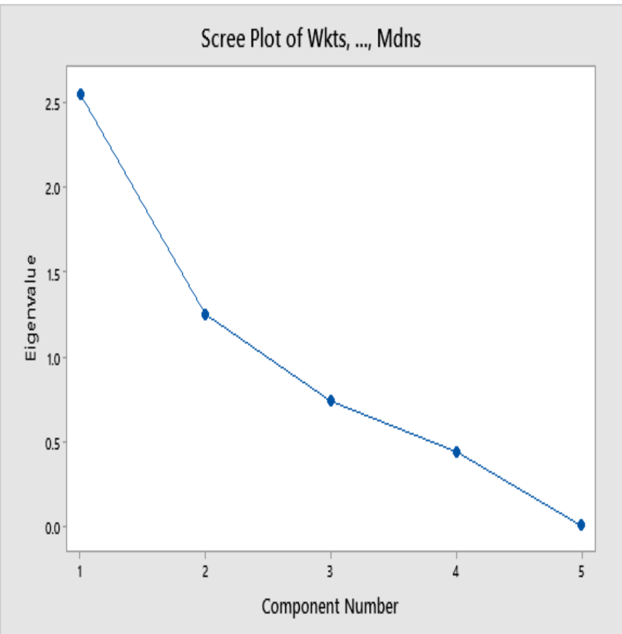

Fig.4. Scree Chart of IPL Bowling

Table 10. Top Ten Bowlers of IPL Ranked by First Principal Component, L1

Ranking	**Player**	**Mat**	**Inn**	**Overs**	**Maiden**	**Wkts**	**Ave**	**Econ**	**SR**	**L1**
1	B Kumar	58	58	223.3	5	71	23.577	7.496	18.870	-9.394
2	JJ Bumrah	60	60	226.6	3	71	22.915	7.180	19.141	-9.013
3	YS Chahal	54	54	191.6	2	65	22.815	7.740	17.686	-6.544
4	Imran Tahir	39	39	147.6	1	55	20.472	7.628	16.101	-3.737
5	Rashid Khan	46	46	182	3	55	21.690	6.554	19.854	-2.135
6	UT Yad	48	48	164.9	1	55	26.054	8.690	17.989	0.647

	av									
7	Sandeep Sharma	50	50	184.4	1	56	25.875	7.857	19.757	0.755
8	MJ McClenaghan	44	44	165.2	1	53	26.735	8.577	18.70	2.282
9	AJ Tye	26	26	99	0	39	21.076	8.303	15.230	4.072
10	K Rabada	18	18	68.2	0	31	17.935	8.152	13.2	4.9238

Bowling Ranking of IPL

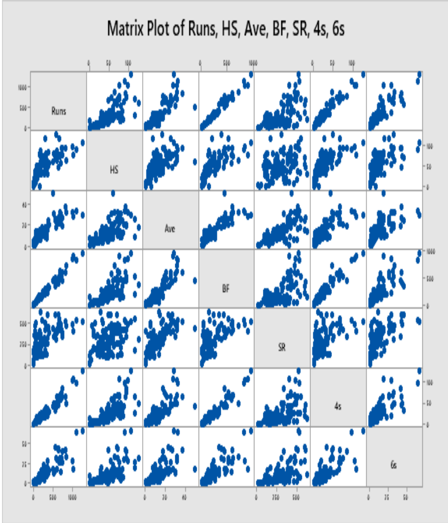

Fig.5. Matrix plots of eight batting parameters of PSL

Bowling lowers the value of L1 depicts better performance and the higher the value of L1 in bowling Analysis depicts poor performance Based on which bowling ranking is generated. Table 10 shows (IPL 2016-2019) lists the top ten bowlers using the first principal component L1.

VII. INTRODUCTION TO PSL (I TO IV)

The Pakistan Super League (PSL) is a Twenty20 cricket league owned by the Pakistan Cricket Board. It is a commercial professional league similar to IPL and BPL, which started in Sep-2015. This league initially had five teams participating. But with the time now, it comprises six teams. Each team is owned by a franchise. PSL on the world calendar is scheduled for February and March of every year. Each team plays double matches. The first four on the point table qualify for the play-offs and the winner of the play-offs will play the final match.

7.1. Result of Batting Performance

It shows that there are noteworthy connections that exist among these measures. Runs and HS We have the following Eigen Values and Eigenvector of the sample co-relation matrix as shown in Table 11. Ordered Eigenvalues and their corresponding percentage total variability are shown in Table 12.

Table 12 represents the ordered Eigenvalues and percentage of total variability attributed to each Eigenvalue. This shows that the first PC explains the most variation that is 78.4%.

Table 13 shows the eigenvector factors for all eight principal components.

Table 11. Sample Correlation Matrix of Batting Performance PSL

Variable	**PC 1**	**PC 2**	**PC 3**	**PC 4**	**PC 5**	**PC 6**	**PC 7**	**PC 8**
RUMS	1	0.744	0.847	0.986	0.616	0.893	0.968	0.888
HS	0.744	1	0.638	0.719	0.325	0.713	0.741	0.675

AVE	0.847	0.638	1	0.857	0.688	0.687	0.783	0.725
BF	0.986	0.719	0.857	1	0.589	0.885	0.959	0.819
SR	0.616	0.325	0.688	0.589	1	0.402	0.533	0.627
50	0.893	0.713	0.687	0.885	0.402	1	0.927	0.735
4s	0.968	0.741	0.783	0.959	0.533	0.927	1	0.812
6s	0.888	0.675	0.725	0.819	0.627	0.735	0.812	1

	4	60	61			32		41
6s	0.356	0.096	-0.128	-0.734	-0.484	0.201	0.050	-0.181

Table 12 Eigen Values and Corresponding and total Variability

Eigen Values	**Total Variability**
6.2738	78.4127
0.8027	10.0325
0.3572	4.464442
0.2761	3.450819
0.1771	2.213473
0.0799	0.998625
0.0299	0.373703
0.0034	0.042495

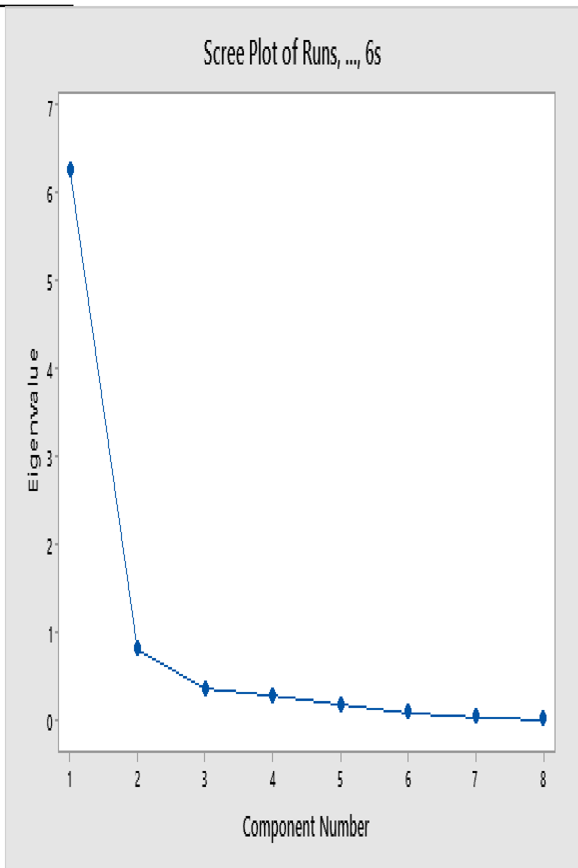

Fig.6. Score Chart of PSL Batting

Table 13 Eigen Values and Eigen Vectors for Sample Correlation Matrix Batting Performance PSL

Variable	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**
Runs	0.395	-0.033	0.147	-0.061	-0.110	-0.279	-0.196	0.830
HS	0.315	-0.396	-0.817	0.054	0.264	-0.054	-0.043	-0.011
Ave	0.351	0.253	-0.131	0.649	-0.484	0.337	0.161	0.001
BF	0.388	-0.047	0.222	0.162	-0.086	-0.456	-0.547	-0.507
SR	0.263	0.797	-0.121	-0.086	0.521	0.008	-0.032	-0.015
50	0.357	-0.324	0.382	-0.011	0.381	0.669	-0.169	-0.016
4s	0.38	-0.1	0.2	0.008	0.163	-0.3	0.776	-0.1

Table 14. Top Batsmen of PSL Ranked by First Principal Component Analysis L1

Ranks	**Player**	**Mat**	**Inns**	**NO**	**Runs**	**HS**	**Ave**	**BF**	**SR**	**100**	**50**	**0**	**4s**	**6s**	**LI**
1	Kamran	47	46	2	1286	107	28.7375	956	523.27	2	9	8	127	67	1136.14

	Akmal														
2	S R Watson	37	37	4	1114	91	33.035	825	534.15	0	7	1	97	65	1003.75
3	Babar Azam	35	34	2	1043	77	27.62	904	407.73	0	9	4	108	18	955.012
4	Ahmed Shehzad	38	36	2	1016	99	31.7325	833	487.95	0	9	3	104	29	948.651
5	Shoaib Malik	39	36	8	849	65	31.63	717	472.83	0	4	2	47	33	800.706
6	Umar Akmal	32	30	5	833	93	37.565	604	518.63	0	7	3	66	42	785.062
7	Mohammad Hafeez	35	33	2	690	77	21.272	605	473.46	0	4	3	69	26	700.711
8	D R Smith	28	25	4	701	73	31.937	611	428.32	0	5	1	67	31	699.365
9	C S Delport	26	25	2	682	117	25.255	519	552.42	1	4	4	66	26	697.796
10	Sarfaraz Ahmed	43	34	10	720	56	30.125	579	496.03	0	3	2	58	14	696.048

Figure 6 represents the ordered Eigen values in scree plot already discuss in above analysis. Figure 6 shows the Scree plot for the batting statistics, in scree plot line bend at 2 is satisfied that we can use as first PC and it explains 78.4 percent of the total variability:

$$L_1 = (0.395 * Runs) + (0.315 * HS) + (0.351 * Ave) + (0.388 * BF) + (0.263 * SR) + (0.357 * 50) + (0.384 * 4s) + (0.356 * 6s)$$

Batting Ranking of PSL

We refer to it as the first principal component. It is the average of PC1-PC8. The higher value of L1 indicates better performance and a lower value of L1 leads to poor performance. Based on the highest and lowest value of L1, we rank the players. Table 14 indicates the top 10 batsmen who played at least 4 matches in the PSL (2016-2019) using the first principal component L1. Here is the top ten PSL Bowler's in the Table 15.

7.2. Results of Bowling Performance

Top ten bowlers of the PSL using measures are mentioned. The analysis is presented in Table 15.

Table 15. Top Ten Bowlers of PSL Ranked by First Principal Component, L1

Ranks	**Player**	**Mat**	**Inns**	**Overs**	**Mdns**	**Runs**	**Wkts**	**Ave**	**Econ**	**SR**	**L1**
1	Wahab Riaz	45	44	166.2	0	1130	65	17.38462	6.799037	15.34154	-7.26759
2	Faheem Ashraf	24	24	84.1	1	632	39	16.20513	7.514863	12.93846	2.303702
3	Mohammad Sami	36	36	126.	1	875	42	20.83333	6.933439	18.02857	6.967331

				2							
4	Usman Shinwari	27	26	89.1	0	758	35	21.65714	8.507295	15.27429	9.677215
5	Mohammad Irfan	56	54	187.2	3	1459	51	28.60784	7.793803	22.02353	10.35648
6	Mohammad Nawaz	43	43	150.4	1	1044	43	24.27907	6.941489	20.98605	10.61421
7	Mohammad Amir	37	36	131.5	4	922	39	23.64103	7.011407	20.23077	10.95539
8	Shahid Afridi	37	36	123.4	5	814	35	23.25714	6.596434	21.15429	12.74607
9	Rahat Ali	18	18	65.7	0	498	25	19.92	7.579909	15.768	13.1071
10	Zafar Gohar	10	9	26.2	0	206	13	15.84615	7.862595	12.09231	13.67437

VIII. CONCLUSION AND FUTURE WORK

Determining a player's performance is an exciting job in any sport. It is especially significant in viable sports like cricket, which is impacted by a player's performance by their runs. We applied the IPL data using PCA. In the future, we plan to propose a cluster-based prediction for a team. What will be the effect on a team's ranking if a team is moved from one region to another? Similar studies can also be utilized to rank other sports as well. Later, the h-index and PageRank algorithms were extended to rank cricket teams. This statistical analysis can be used for football leagues using different principal component analysis.

REFERENCES

1. KAMAL, S.A., et al., *BOZOK INTERNATIONAL JOURNAL OF SPORT SCIENCES.*
2. Sharma, S.K., R.G. Amin, and S.J.I.J.o.P.A.i.S. Gattoufi, *Choosing the best Twenty20 cricket batsmen using ordered weighted averaging.* 2012. **12**(3): p. 614-628.
3. Hotelling, H.J.J.E.P., *Analysis of a complex of statistical variables with principal components.* 1933. **24**: p. 498-520.
4. Barr, G. and B.J.J.o.t.O.R.S. Kantor, *A criterion for comparing and selecting batsmen in limited overs cricket.* 2004. **55**(12): p. 1266-1274.
5. Swartz, T.B., et al., *Optimal batting orders in one-day cricket.* 2006. **33**(7): p. 1939-1950.
6. Swartz, T.B., P.S. Gill, and S.J.C.J.o.S. Muthukumarana, *Modelling and simulation for one-day cricket.* 2009. **37**(2): p. 143-160.
7. Preston, I. and J.J.J.o.t.R.S.S.S.D. Thomas, *Batting strategy in limited overs cricket.* 2000. **49**(1): p. 95-106.
8. Van Staden, P.J., *Comparison of cricketers' bowling and batting performances using graphical displays.* 2009.
9. Shah, S., P.J. Hazarika, and J.J.I.J.o.A.R.i.C.S. Hazarika, *A study on performance of cricket players using factor analysis approach.* 2017. **8**(3): p. 656-660.
10. Motipalle, R. and S.V.L. Kumar, *A multivariate statistical approach for ranking the best batsmen in test cricket.* 2020.
11. Shanto, S.I. and N.J.I.J.o.P.A.i.S. Awan, *A sequential principal component-based algorithm for optimal lineup and batting order selection in one day international cricket for Bangladesh.* 2019. **19**(4): p. 567-583.
12. Perera, H., et al., *Optimal lineups in Twenty20 cricket.* 2016. **86**(14): p. 2888-2900.
13. Shah, P. and M.J.I.J.P.N.P.E. Patel, *Ranking the cricket captains using principal component analysis.* 2018. **3**(2): p. 477-483.
14. Bailey, M., S.R.J.J.o.s.s. Clarke, and medicine, *Predicting the match outcome in one day international cricket matches, while the game is in progress.* 2006. **5**(4): p. 480.
15. Mustafa, R.U., et al., *Predicting the cricket match outcome using crowd opinions on social networks: A comparative study of machine learning methods.* 2017. **30**(1): p. 63-76.
16. Pearson, K. and L.O.J.J.S. Lines, *Planes of Closest Fit to Systems of Points in Space, London Edinburgh Dublin Philos. Mag.* 1901. **2**(11): p. 559-572.
17. Hotelling, H.J.J.o.e.p., *Analysis of a complex of statistical variables into principal components.* 1933. **24**(6): p. 417.
18. Hladnik, A.J.I.C.o.G.E. and Research, *Image compression and face recognition: Two image processing applications of principal component analysis.* 2013. **6**(4): p. 56-61.
19. Dawkins, B.J.T.A.S., *Multivariate analysis of national track records.* 1989. **43**(2): p. 110-115.
20. Watnik, M. and R.A.J.J.o.S.E. Levine, *NFL Y2K PCA.* 2001. **9**(3).
21. Manage, A.B. and S.M.J.J.o.S.E. Scariano, *An introductory application of principal components to cricket data.* 2013. **21**(3).
22. Kaiser, H.F.J.E. and p. measurement, *The application of electronic computers to factor analysis.* 1960. **20**(1): p. 141-151.