SVKM'S NMIMS Deemed to be UNIVERSITY

NILKAMAL SCHOOL OF MATHEMATICS, APPLIED STATISTICS & ANALYTICS

STATISTICAL ANALYSIS OF INDIAN CRICKET TEAM IN ONE DAY INTERNATIONAL

INTRODUCTION

In cricket, India has become a dominant force globally. This report presents statistics which highlight a story on India's cricketing journey.

We aim at revealing the trends, highlights, and the pattern of the team's achievement in ODI cricket. We'll discuss how numbers can convey strengths, weaknesses, top players, and the group's experience as a whole. We will examine different elements of the team's statistical environment ranging from batting averages to bowling figures ,win-loss ratios and many more.

A One Day International (ODI) is a form of limited overs cricket, played between two teams with international status, in which each team faces a fixed number of overs, currently 50.[1]

CRICKET TERMINOLOGIES

BATTING

- *Batting Average* - The total number of runs they have scored divided by the number of times they have been out.
- *Balls Faced* - The total number of balls received, including no-balls but not including wides.
- *Batting Strike Rate*-The average number of runs scored per 100 balls faced.
- *Inning*-An innings is one of the divisions of a cricket match during which one team takes its turn to bat.

BOWLING

- *Bowling Average*-The ratio of number of runs conceded per wickets taken
- *Economy Rate*-Number of runs bowler is conceding per overs bowled.
- *Bowling strike rate*-The average number of balls bowled per wicket taken.
- *Maidens*-An over in which no wides or no balls are bowled, and no runs are scored off the bat.

OBJECTIVES

- To compare team's Batting Performances of 2019 and 2022.
- To test equality of means of Economy Rate of spinners and pacers.
- To test Rohit Sharma's Captaincy.
- To test the independence of Venue and Outcome of the match.
- To test the independence of toss winning and outcome of the match.
- To test the variances of batting performances and Economy Rates.
- To select ideal Indian cricket team for ODI

DATA

We have used secondary data for our research
Source:
https://stats.espncricinfo.com/ci/engine/stats/index.html

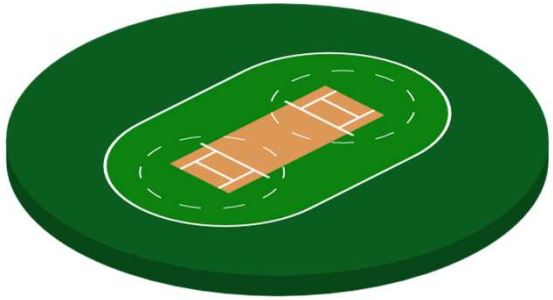

STATISTICAL
ANALYSIS

1)CHI SQUARE TEST OF INDEPENDENCE OF VENUE AND OUTCOME OF THE MATCH

$H_0 = Venue\ and\ Outcome\ of\ a\ match\ are\ independent$

$H_1 = Venue\ and\ Outcome\ of\ a\ match\ are\ not\ independent$

$\alpha = 0.05$



ASSUMPTIONS:

1) $N \geq 50$

2)Both variables are categorical

1)CHI SQUARE TEST OF INDEPENDENCE OF VENUE AND OUTCOME OF THE MATCH

$H_0 = Venue\ and\ Outcome\ of\ a\ match\ are\ independent$

$H_1 = Venue\ and\ Outcome\ of\ a\ match\ are\ not\ independent$

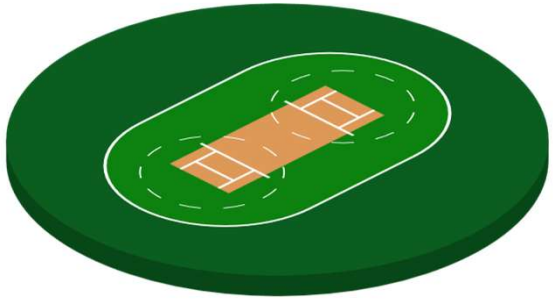
$\alpha = 0.05$

$\chi^2 = 0.66466 \quad < \quad \chi^2_{1,0.05} = 3.84$

$therefore,\ we\ accept\ H_0$

CONCLUSION:

Venue and Outcome of a match are independent.ICT has excellent adaptibility.



Formula

$$\therefore \chi^2 = \Sigma\left(\frac{O_{ij}^2}{E_{ij}}\right) - N$$

$Where,\ Oi = Observed\ Value$

$Ei = Expected\ Value$

2)TO TEST THE INDEPENDENCE OF TOSS WINNING AND OUTCOME OF THE MATCH.

$H_0 = Winning\ the\ toss\ and\ Outcome\ of\ a\ match\ are\ independent$

$H_1 = Winning\ the\ toss\ and\ Outcome\ of\ a\ match\ are\ not\ independent$

$\alpha = 0.05$

ASSUMPTIONS:

1) $N \geq 50$

2)Both variables are categorical

Formula:

$$\chi^2 = \Sigma\left(\frac{O_{ij}^2}{E_{ij}}\right) - N$$

$Where,\ Oi = Observed\ Value$

$Ei = Expected\ Value$

2)TO TEST THE INDEPENDENCE OF TOSS WINNING AND OUTCOME OF THE MATCH.

$H_0 = Winning\ the\ toss\ and\ Outcome\ of\ a\ match\ are\ independent$

$H_1 = Winning\ the\ toss\ and\ Outcome\ of\ a\ match\ are\ not\ independent$

$\alpha = 0.05$

$$\chi^2 = 0.9981 \quad < \quad \chi^2_{1,0.05} = 3.84$$

$therefore,\ we\ accept\ H_0$

CONCLUSION:

Winning a toss and the outcome of a match are independent.

Formula:

$$\chi^2 = \Sigma\left(\frac{O_{ij}^2}{E_{ij}}\right) - N$$

$Where,\ Oi = Observed\ Value$

$Ei = Expected\ Value$

3) FITTING OF POISSON DISTRIBUTION

$H_0 = Poisson\ Distribution\ is\ a\ good\ fit.$

$H_1 = Poisson\ Distribution\ is\ not\ a\ good\ fit$

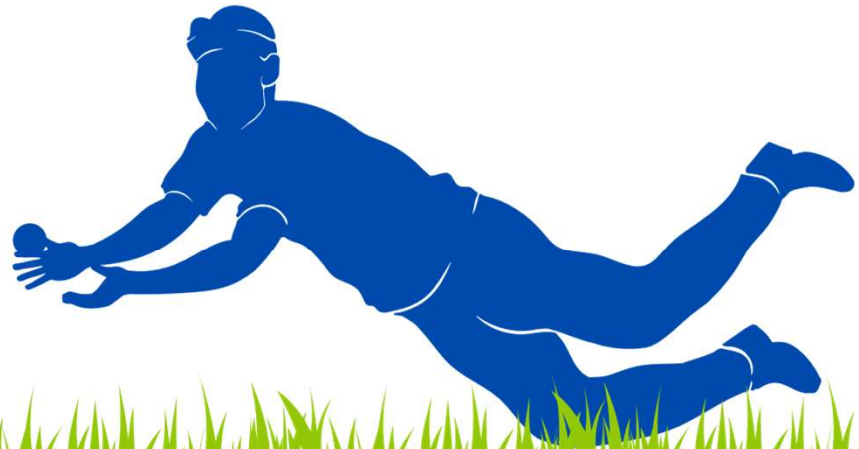
$\alpha = 0.05$

ASSUMPTIONS:

1) $N \geq 50$

*2)*ΣOi = Σ*Ei* =*N*

3)$Each\ Ei \geq 5$



3) FITTING OF POISSON DISTRIBUTION

$$H_0 = Poisson\ Distribution\ is\ a\ good\ fit.$$

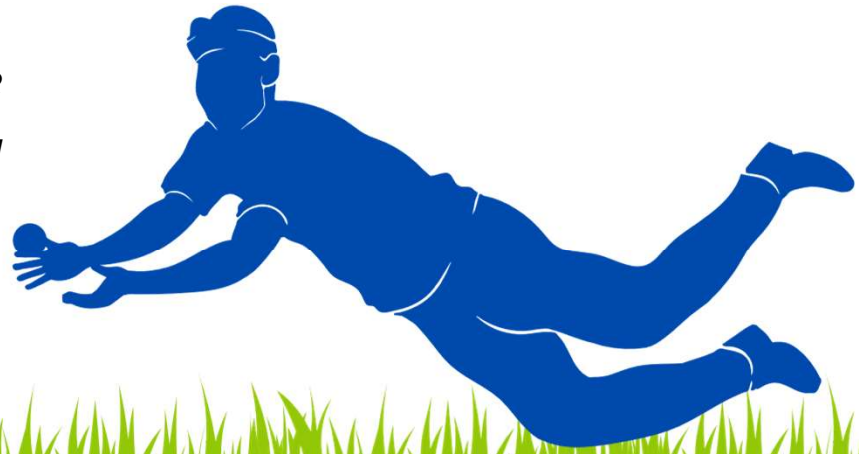
$$H_1 = Poisson\ Distribution\ is\ not\ a\ good\ fit$$

$$\alpha = 0.05$$

$$p - value = 0.475$$

CONCLUSION:

Since P-value is greater than 0.05 therefore we accept null hypothesis Poisson Distribution is a good fit

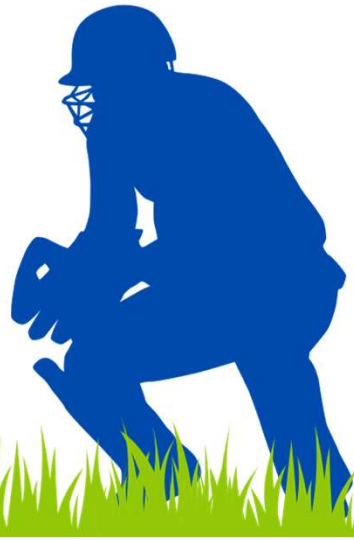


CODE

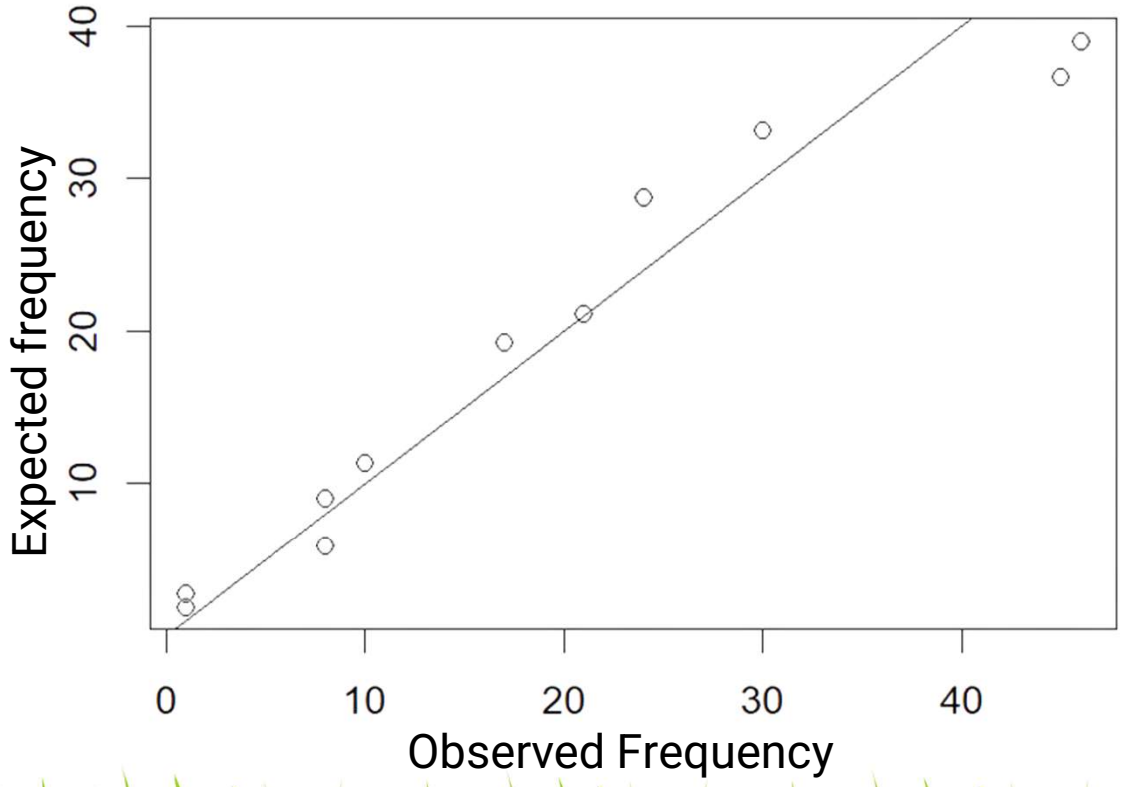
```
library("vcd")

lambda<-goodfit(catchestaken,type="poisson",method="ML")$par

observed <-goodfit(catchestaken,type="poisson",method="ML")$observed

expected <-goodfit(catchestaken,type="poisson",method="ML")$fitted

chi_square_stat <- sum((observed-expected)^2/expected)

df <- length(observed) - 1-1

p_value <- 1 - pchisq(chi_square_stat, df)

print(paste("Chi-Square Statistic:", chi_square_stat))
print(paste("Degrees of Freedom:", df))
print(paste("P-value:", p_value))
print(lambda)

qqplot(observed,fitted,abline(0,1),xlab="Observed Frequencies",ylab="Expected Frequencies")
```

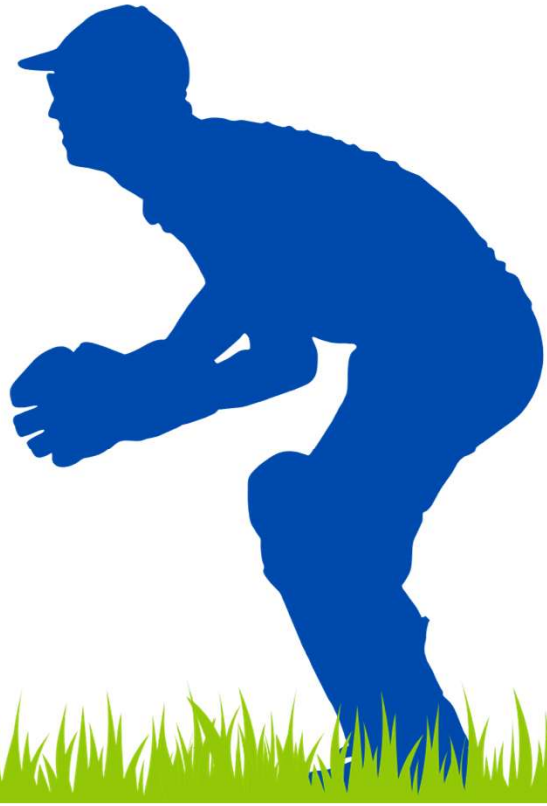
OUTPUT

```
[1] "Chi-Square Statistic: 8.59671597768546"
> print(paste("Degrees of Freedom:", df))
[1] "Degrees of Freedom: 9"
> print(paste("P-value:", p_value))
[1] "P-value: 0.475301529806837"
> print(lambda)
$lambda
[1] 4.701422
```



Q-Q PLOT OF OBSERVED AND EXPECTED FREQUENCY OF CATCHES TAKEN





4) TESTING CAPTAINCY OF ROHIT SHARMA FOR THE YEARS 2019-2023 USING BINOMIAL TEST OF SINGLE POPULATION PROPORTION.

P : *Population Proportion of India winning the match*

p_0 :*Sample Proportion of India winning the match*

ASSUMPTIONS:

1)The outcome of each observation is binary
2)Each observation or trial is independent of others.
3)Random sampling
4)Fixed number of trials

4) TESTING CAPTAINCY OF ROHIT SHARMA FOR THE YEARS 2019-2023 USING BINOMIAL TEST OF SINGLE POPULATION PROPORTION.

P : *Population Proportion of India winning the match*

p_0 :*Sample Proportion of India winning the match*

$$H_0 = P \geq 0.80 \quad , \quad H_1 = P < 0.80$$

α =0.05 $P_0 = 0.69$ $p - value = 0.1313$ $n = 26$

Formula:

$$p - value = P(X <= x \mid p = P_0) = \sum_{i=0}^{x} C(n,k) P_0{}^{i} (1 - P_0)^{(n-i)}$$

Conclusion:

$$Since\ p\ value = 0.1313 > 0.05,\ we\ accept\ H_0$$

5)COMPARING VARIANCES OF TWO INDEPENDENT POPULATIONS OF ECONOMY RATES OF FAST BOWLERS AND SPINNERS

$H_0: \sigma_1^2 = \sigma_2^2$ $\quad$ $H_1: \sigma_1^2 \neq \sigma_2^2$

$\alpha = 0.05$

ASSUMPTIONS:

1) The populations follow normal distribution
2)The populations are independent
3)Greater of the 2 variances is to be taken in the numerator and n1 corresponds to the greater variance.

5)COMPARING VARIANCES OF TWO INDEPENDENT POPULATIONS OF ECONOMY RATES OF FAST BOWLERS AND SPINNERS

$H_0: \sigma_1^2 = \sigma_2^2$ $\quad$ $H_1: \sigma_1^2 \neq \sigma_2^2$

$\alpha = 0.05$

p value=0.06127

0.06127>0.05

$therefore,\ we\ accept\ H_0$

CONCLUSION:

Variances of Economy Rates of bowlers is same

Formula:

$$F = \frac{S_1^2}{S_2^2}$$

$S_1^2 > S_2^2$

CODE

```
var.test(pacers,spinners)
```

OUTPUT

```
        F test to compare two variances

data:  spinners and pacers
F = 2.5238, num df = 16, denom df = 18, p-value = 0.06127
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.9558452 6.8570840
sample estimates:
ratio of variances
          2.523767
```

6) TESTING FOR EQUALITY OF MEAN OF ECONOMY RATE OF SPIN BOWLER AND FAST BOWLER.

$H_0 = mean\ of\ Economy\ Rate\ of\ spin\ bowler\ and\ fast\ bowler\ are\ same.$

$H_1 = mean\ of\ Economy\ Rate\ of\ spin\ bowler\ and\ fast\ bowler\ are\ not\ same.$

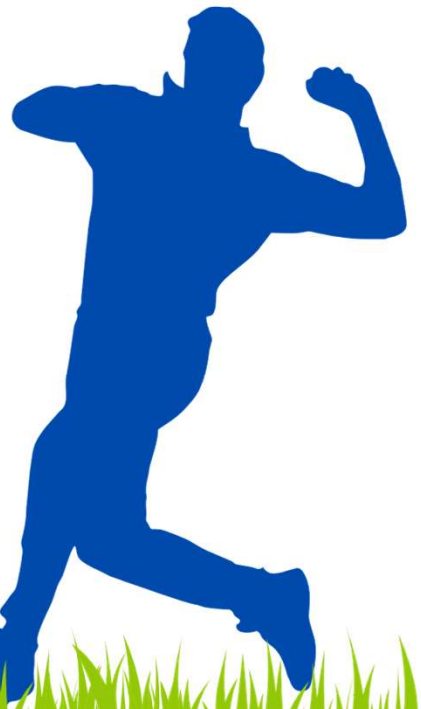
$\alpha = 0.05$

ASSUMPTIONS:

1)The two samples are independent

2)Both samples have been drawn from normal population

3)Population variances are equal and unknown



6) TESTING FOR EQUALITY OF MEAN OF ECONOMY RATE OF SPIN BOWLER AND FAST BOWLER.

$H_0 = mean\ of\ Economy\ Rate\ of\ spin\ bowler\ and\ fast\ bowler\ are\ same.$

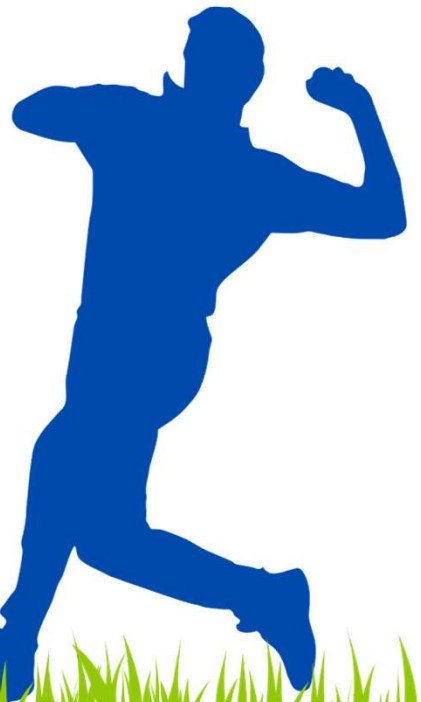
$H_1 = mean\ of\ Economy\ Rate\ of\ spin\ bowler\ and\ fast\ bowler\ are\ not\ same.$

$\alpha = 0.05$

Formula:

$$t = \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where, $S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}$



CODES

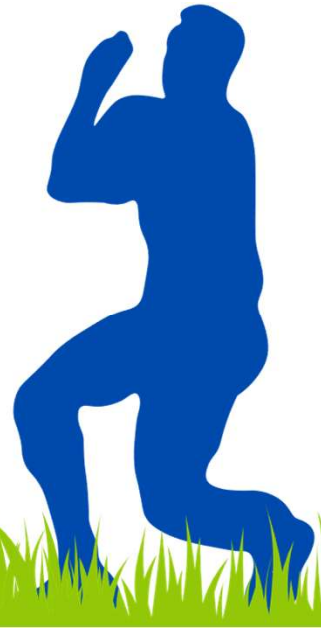
```
dataoffast<-c(6.75,6.02,5.8,4.25,5.58,5.74,3.2,5.16,
              7.4,5.32,6.1,4.62,4.6,6.87,8.68,6.18,7.15,6.54,5.6);
dataoffast;

dataofspin<-c(4.82,4.76,5.3,6.12,5.86,11,5.53,10,
              3.33,6.05,5.31,8.62,5.4,4.8,7.16,4.86,5.76)
dataofspin;

t_test_result <- t.test(dataoffast, dataofspin,var.equal= T)
t_test_result
```

OUTPUT

```
        Two Sample t-test

data:  dataoffast and dataofspin
t = -0.52259, df = 34, p-value = 0.6047
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.3985337  0.8263974
sample estimates:
mean of x mean of y
 5.871579  6.157647
```

CONCLUSION:

The means of Economy Rate of spin bowler and fast bowler are equal. Hence,we can conclude that the performance of both spinners and fast bowlers is equally good for the team.



7)COMPARING VARIANCES OF TWO INDEPENDENT POPULATIONS OF BATTING PERFORMANCE PLAYERS IN 2019 TO 2022

$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$

$\alpha = 0.05$ $n_1 = 7$ $n_2 = 13$

ASSUMPTIONS:

1) The populations follow normal distribution
2)The populations are independent
3)Greater of the 2 variances is to be taken in the numerator and n1 corresponds to the greater variance.

7)COMPARING VARIANCES OF TWO INDEPENDENT POPULATIONS OF BATTING PERFORMANCE PLAYERS IN 2019 TO 2022

$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$

$\alpha = 0.05$ $n_1 = 7$ $n_2 = 13$

P value=0.04466

0.04466<0.05

$therefore, we\ reject\ H_0$

CONCLUSION:

Variance of both the population of batting performances are not same

Formula:

$$F = \frac{S_1^2}{S_2^2}$$

$S_1^2 > S_2^2$

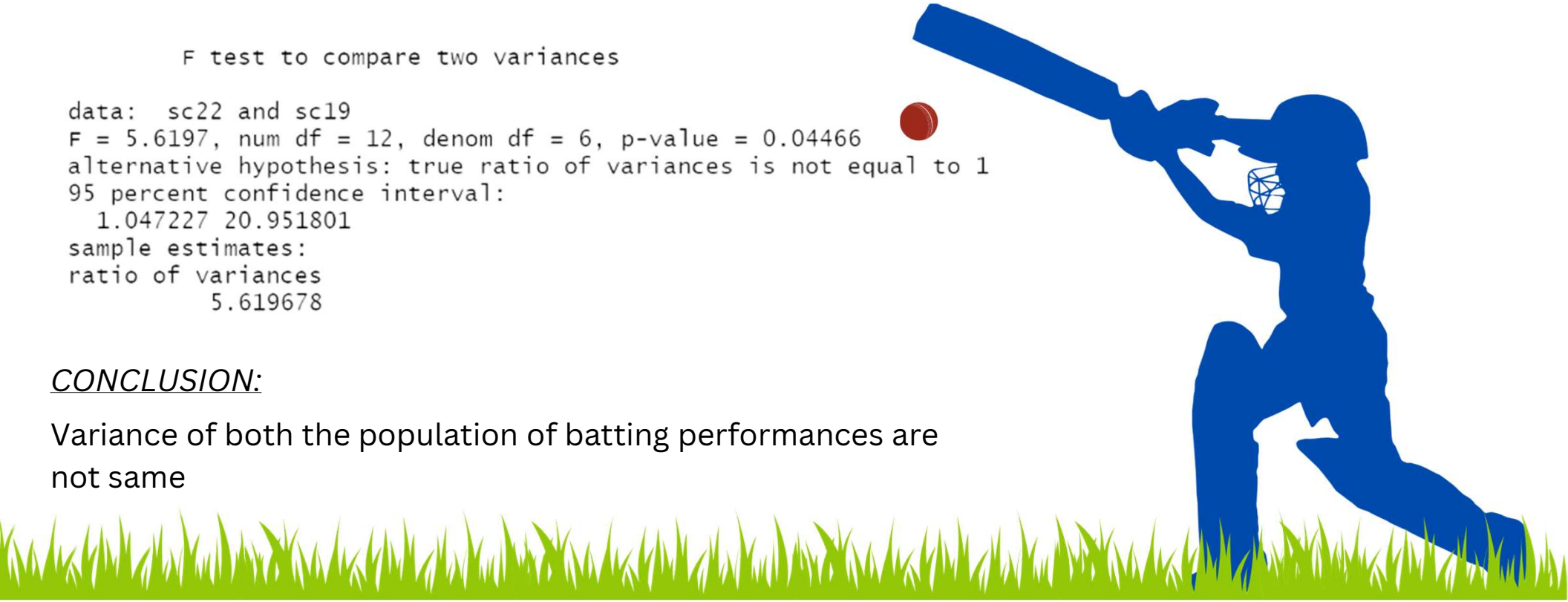
$n_1 = 7$ $n_2 = 13$

CODE

```
var.test(sc22,sc19)
```

OUTPUT

```
        F test to compare two variances

data:  sc22 and sc19
F = 5.6197, num df = 12, denom df = 6, p-value = 0.04466
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  1.047227 20.951801
sample estimates:
ratio of variances
          5.619678
```



CONCLUSION:

Variance of both the population of batting performances are not same

8) COMPARING TEAM BATTING PERFORMANCES OF 2019 AND 2022 USING T-TEST FOR 2 POPULATION MEAN.

$$H_0 = Mean\ batting\ performance\ of\ both\ years\ is\ same$$

$$H_1 = Mean\ batting\ performance\ of\ both\ years\ is\ not\ same$$

ASSUMPTIONS:

1)The two samples are independent
2)Both samples have been drawn from normal population
3)The variances are unequal and unknown

8) COMPARING TEAM BATTING PERFORMANCES OF 2019 AND 2022 USING T-TEST FOR 2 POPULATION MEAN.

H_0 = *Mean batting performance of both years is same*

H_1 = *Mean batting performance of both years is not same*

α =0.05

$\bar{x}$ = *sample mean of matches played in* 2019 = 313

$\bar{y}$ = *sample mean of matches played in* 2022 = 280.3077

$p - value = 0.3642$

$p - value = 0.3642 \geq 0.05$

therefore, we accept H_0

Formula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$n_1 = 7 \quad n_2 = 13$

CODES

```
t.test(sc19,sc22,var.equal = F)
```

CONCLUSION:

Mean batting performance of both years is same.
ICT is consistent.

OUTPUT

```
        Welch Two Sample t-test

data:  sc19 and sc22
t = 0.93164, df = 17.434, p-value = 0.3642
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -41.20302 106.58763
sample estimates:
mean of x mean of y
 313.0000  280.3077
```

SELECTION OF AN IDEAL TEAM

SELECTION OF BATTERS- ROUND 1

- From ESPN website, the data of batters of Indian Cricket Team was taken.It consisted of 49 players.
- We considered 3 parameters as part of the first round of selection.The parameters were Strike Rate,Avg,Highest score,
- The top 20 batters in each of the above parameters were chosen in that parameter.The batters that made it to the top 20 of every parameter were considered as potential batters.
- This resulted in having a total of **13 potential batters.**

SELECTION OF BATTERS- ROUND 2

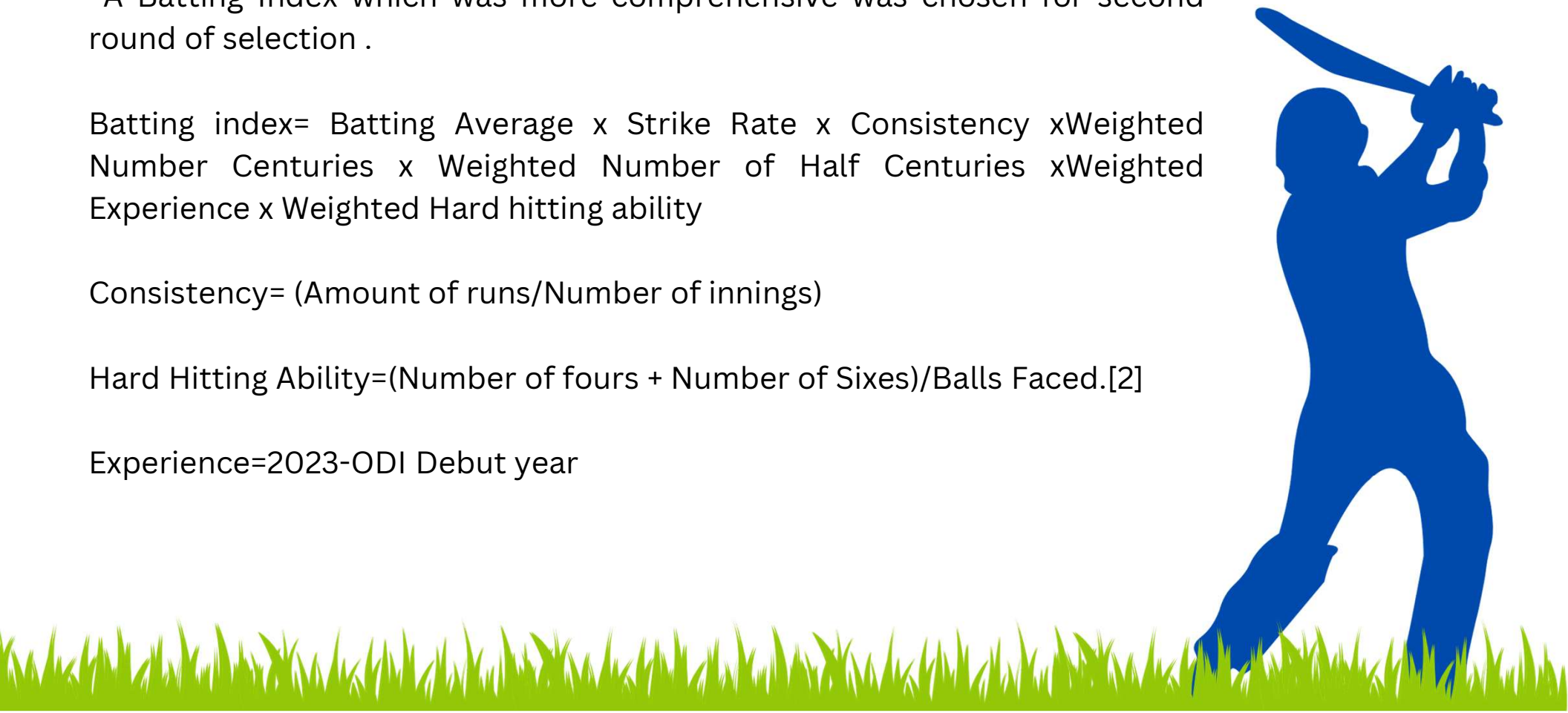
A Batting Index which was more comprehensive was chosen for second round of selection .

Batting index= Batting Average x Strike Rate x Consistency xWeighted Number Centuries x Weighted Number of Half Centuries xWeighted Experience x Weighted Hard hitting ability

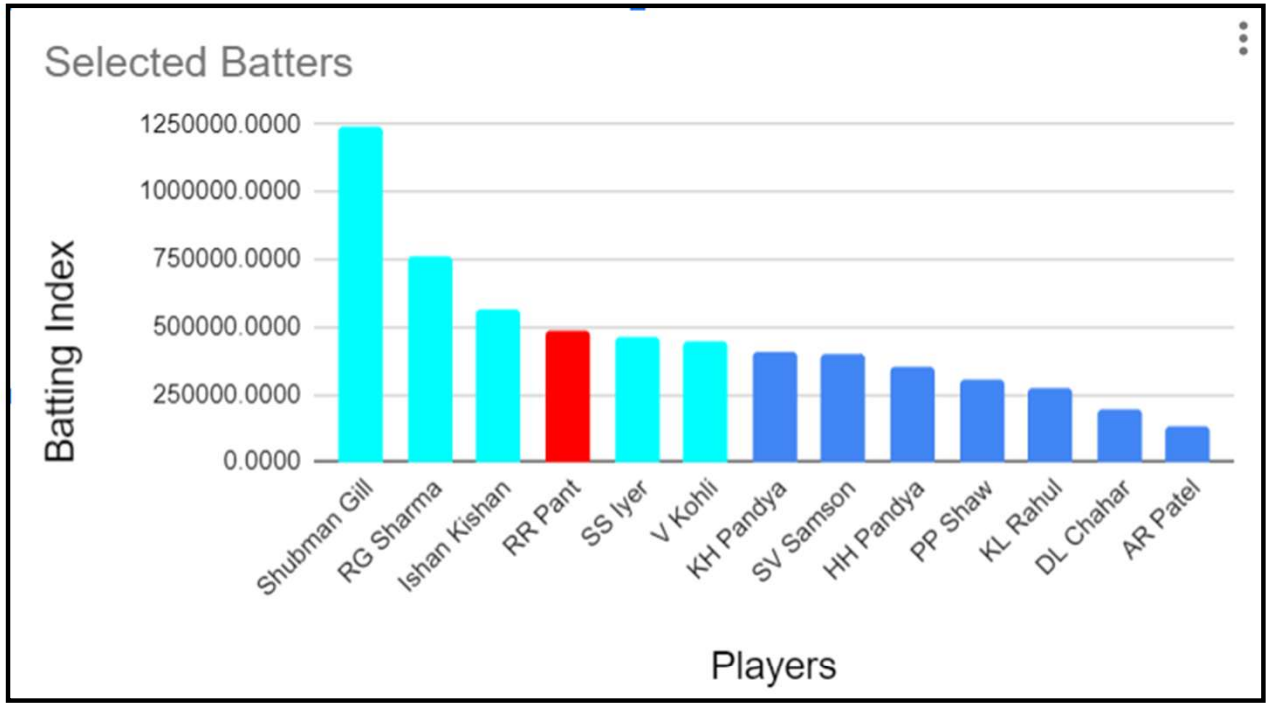
Consistency= (Amount of runs/Number of innings)

Hard Hitting Ability=(Number of fours + Number of Sixes)/Balls Faced.[2]

Experience=2023-ODI Debut year



SECOND ROUND OF SELECTION OF POTENTIAL BATTERS



Please note that RR pant is eliminated from the second selection procedure because he has been injured since December 2022.He is not fit to be a part of the current ODI team.

BATTERS

ROHIT SHARMA
TOP ORDER BATTER

VIRAT KOHLI
TOP ORDER BATTER

SHREYAS IYER
TOP ORDER BATTER

SHUBMAN GILL
OPENING BATTER

ISHAN KISHAN
WICKETKEEPER BATTER

SELECTION OF BOWLERS- ROUND 1

- The data regarding the bowlers was taken from ESPN. It consisted of 49 players.
- We considered 3 parameters as part of the 1st round of selection.The parameters considered were economy, bowling average and strike rate.
- The top 20 bowlers in each of the above parameters were chosen in that parameter. The bowlers that made it to the top 20 of every parameter were considered as potential bowlers.
- This resulted in having a total of **13 potential bowlers.**

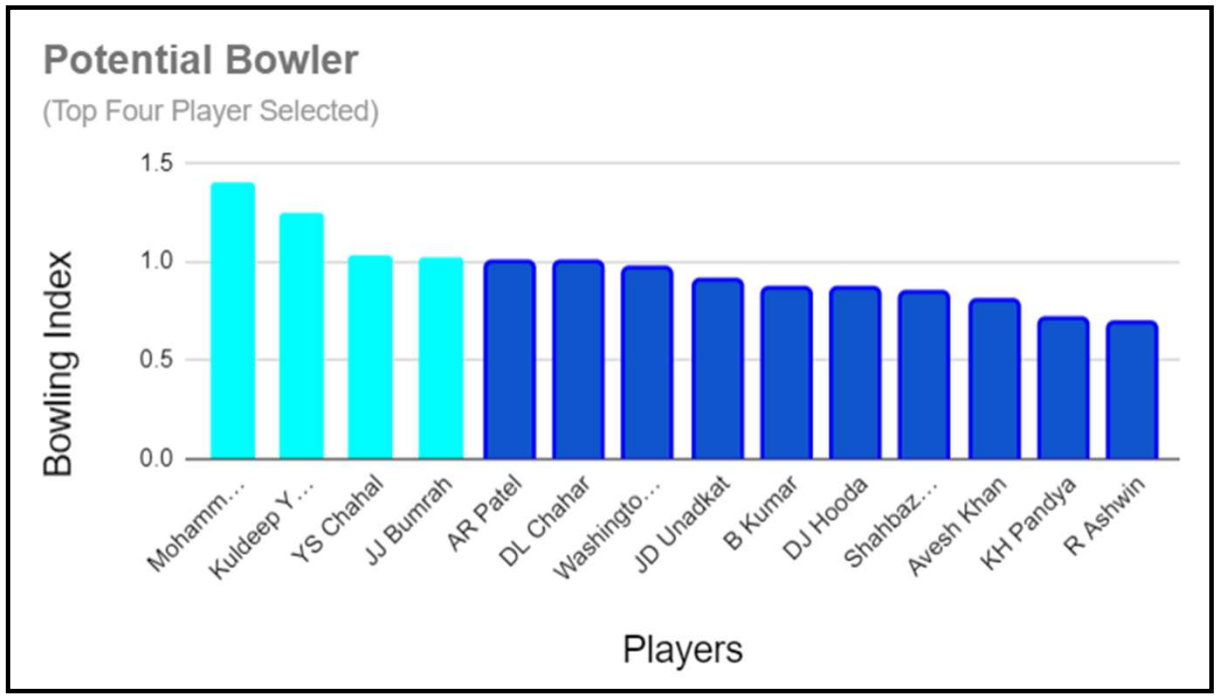
SELECTION OF BOWLERS- ROUND 2

A Bowling Index which was more comprehensive was chosen.

BOWLING INDEX:

$$\frac{\text{(WEIGHTED MAIDEN X WEIGHTED WICKETS)}}{\text{(WEIGHTED ECONOMY X WEIGHTED STRIKE RATE X WEIGHTED BOWLING AVERAGE)}}$$

SECOND ROUND OF SELECTION OF POTENTIAL BOWLERS

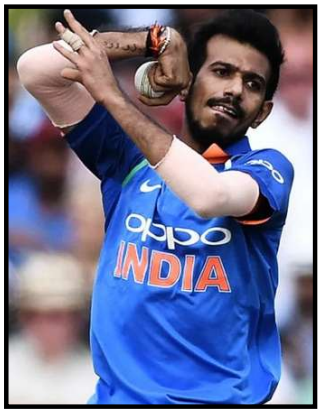


BOWLERS

JASPRIT BUMRAH
RIGH ARM FAST

MOHAMMED SIRAJ
RIGHT ARM FAST

KULDEEP YADAV
LEFT ARM WRIST SPIN



YUZVENDRA CHAHAL
LEGBREAK GOOGLY

SELECTION OF ALL ROUNDERS- ROUND 1

- The data of All rounders was taken from ESPN.ICT had a total of 16 Batting all rounders and 12 Bowling all rounders.
- The players who were common in both batting and bowling all rounders were chosen.

SELECTION OF ALL ROUNDERS- ROUND 2

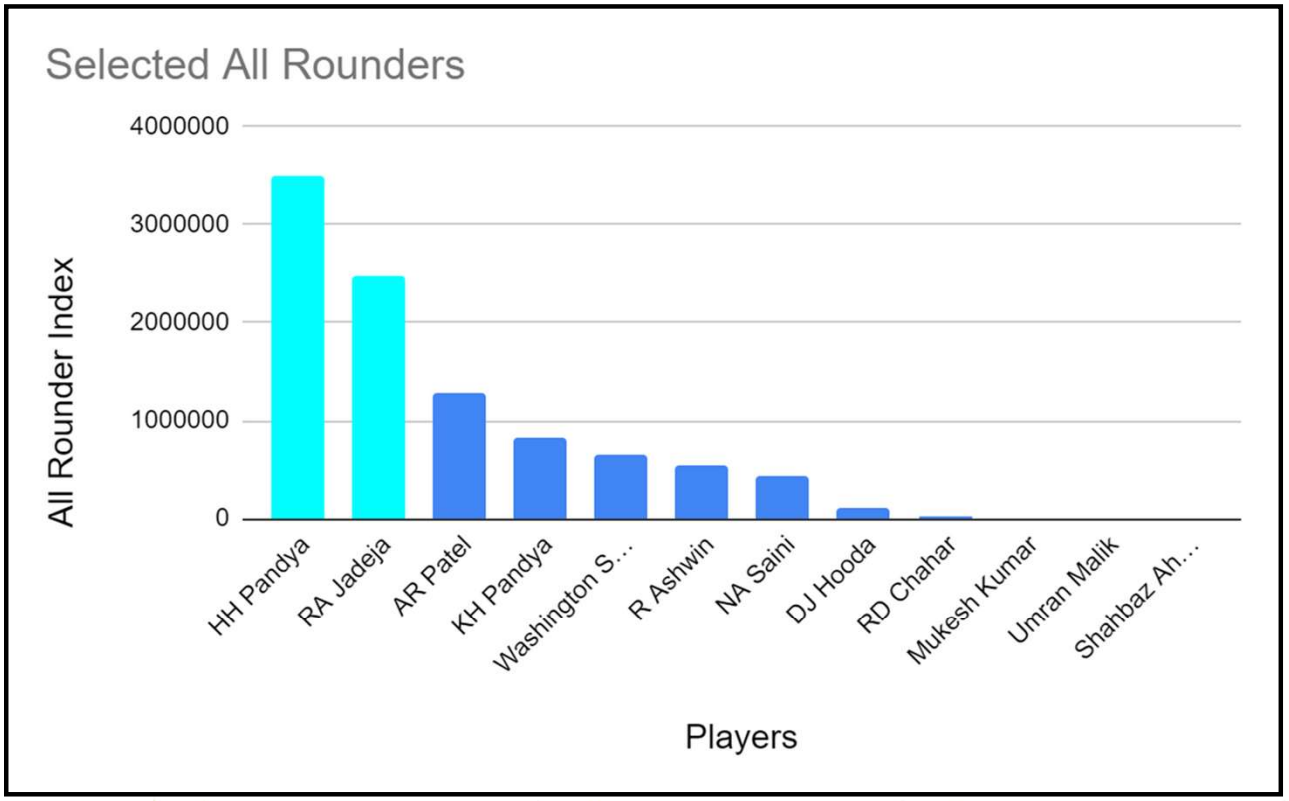
All rounder index was calculated for these players.
All rounder index= Batting Index x Balling Index

Batting Index= Batting Average x Strike Rate x Consistency x Weighted Number of Half Centuries x Weighted Experience x Weighted Hard hitting ability

Bowling index:

$$\frac{\text{(weighted maiden x weighted wickets)}}{\text{(weighted economy x weighted strike rate x weighted bowling average)}}$$

SECOND ROUND OF SELECTION OF POTENTIAL ALL ROUNDERS



ALL ROUNDERS

HARDIK PANDYA

RAVINDRA JADEJA


ICC MEN'S
CRICKET WORLD CUP
INDIA 2023
INDIA

FUTURE SCOPE

- Future research can explore how specific player statistics impact a team's success in different game formats.
- Comparing Indian Cricket Team's stats with those of other top cricketing nations can identify areas for improvement and strengths.
- Advanced machine learning algorithms can help develop predictive models to guide team selection, strategy formulation, and performance evaluation.
- Incorporating more years of data and additional parameters, such as fielding statistics, can provide a comprehensive view of player contributions.
- These extensions can enhance sports analytics, particularly in cricket.

CONCLUSION

The study found that factors such as venue, toss outcome, and match outcome are independent of each other. The team's performance is consistent.

A proposed team selection based on calculated indexes resulted in a balanced team with a mix of batters, bowlers, spinners, and all-rounders. The study highlights the importance of statistical analysis in sports and demonstrates how data-driven insights can contribute to strategic decision-making in cricket.

Sets a precedent for how data-driven decision-making can revolutionize the world of sports. Here, every run, every wicket, and every catch is a story told by data.

REFERENCES

[1] WIKIPEDIA CONTRIBUTORS, “ONE DAY INTERNATIONAL,” WIKIPEDIA, OCT. 29, 2023. HTTPS://EN.WIKIPEDIA.ORG/WIKI/ONE_DAY_INTERNATIONAL

[2] PREDICTING OPTIMAL CRICKET TEAM USING DATA ANALYSIS RESEARCH PAPER

Thank you!