Guidance on Students' Project

2021 (Pandemic Year):

Sl. No.	Year	Name of the student	Name of Supervisor	Title of the project
1	2021	Ms. Priyanka Choudhury	Dr. N. Ahmed	Analyzing the Consistency of the 10 randomly selected out of the Top 20 in T-20 ICC ranked batsmen from 25th August, 2018 to 25th August, 2021 using measures of Central Tendency and measures of dispersion.
2	2021	Mr. Navajit Chanda	Dr. N. Ahmed	Testing of consistency by SRS amongst the 10 randomly selected batsmen from Top 20 ICC Men's Ranking Batsmen, 2021.
3	2021	Ms. Panchatapa Purkayastha	Dr. N. Ahmed	A Statistical Analysis of the Most Consistency batswoman in ODI of the top 10 ICC Men's Ranked cricketers, 2021.
4	2021	Mr. Surajit Dutta	Ms. K. Borah	A Study on the Nutritional Status and Prevalence of Anaemia among women under Reproductive Age (15-49).
5	2021	Ms. Jasinta Lakra	Ms. S. Amrin	The use of Tobacco and Alcohol Consumption among the Adults based on the States of India: An analysis based on NFH- 5 Phase -1 States/UTS Factsheets.
6	2021	Ms. Rumpi Choudhury	Ms. S. Amrin	High Blood Sugar Levels in India: An insight into NFHS-5 Data (2019-2020).

A Sample of Student's Project Work 2021

Department of Statistics,

DHSK College

Analysing the consistency of the ten randomly selected batsmen from the top twenty T-20, ICC ranked batsmen using the measures central tendency and dispersion

A

Project Report Submitted

By

Priyanka Choudhury

In partial fulfillment for the award of the degree of Bachelor of Science With major in Statistics.

Under the guidance of

Dr. Nazimuddin Ahmed

Department of Statistics
D.H.S.K. College
Affiliated to Dibrugarh University, Dibrugarh
2021

Dr. Nazimuddin Ahmed, M.Sc., M. Phil., Ph.D.
Head, Department of Statistics
D.H.S.K. College
Dibrugarh- 786001
Assam, India
E mail: mdnahmed@yahoo.com

Certificate

This is to certify that the project report entitled *"**Analysing the consistency of the 10 randomly selected out of the top 20 ICC ranked batsmen from 25th August, 2018 to 25th August, 2021***", prepared by Priyanka Choudhury, a student of B.Sc. 6th Semester, D.H.S.K. College, Dibrugarh to Dibrugarh University, Dibrugarh, for the award of the bachelor degree in Statistics is a record of bonafide project work carried out by the candidate under my supervision and guidance. The field report has reached the standard fulfilling the requirements of the regulation for the award of the degree. The results embodied in this project report have not been submitted to any other University or Institute for the award of any degree or diploma.

Date:
Place: Dibrugarh (Nazimuddin Ahmed)

ACKNOWLEDGMENT

It is my pleasure and privilege to express my deep sense of gratitude to, Dr. Nazimuddin Ahmed, Asst. Prof. of D.H.S.K. College, who initiated me into this interesting field of research, has extended his valuable suggestions and cooperation on handling various computer operations during the period of project report and whose esteemed guidance has enabled me to present this project report.

I also thank all the other faculty members of the Department of Statistics, D.H.S.K. College, Dibrugarh, Ms. Krishna Borah, and Ms. Sahista Amreen for their constant encouragement and moral support in shaping up my academic engagement with the discipline of statistics. This dissertation is somehow a partial reflection of their contribution towards my academic growth over the last three years. I, therefore express a deep sense of gratitude and heartfelt devotion to them.

Lastly, I thank to all others, who directly or indirectly helped me in bringing out the thesis in the present form.

(Priyanka Choudhury)

TABLE OF CONTENTS

GLOSSARY	**I**
LIST OF FIGURES	**II**
LIST OF TABLES	**III**
CHAPTER 1: INTRODUCTION	**1**
1.1Sports Analytics	**2**
1.2Introduction to Cricket	**3**
1.3Importance of Statistical Data Analysis in Cricket	**6**
1.4A Brief Review of Literature	**7**
1.5Some Relevant Definitions Used in the Study	**8**
1.6Problem and Purpose of the Project	**11**
1.7Chapter-wise Summary	**12**
CHAPTER 2: DATA COLLECTION	**13**
2.1 Research Methodology	**14**
2.2 Objectives of the Present Study	**14**
2.3 Data Collection	**15**
2.4 Collection of Statistical Data	**16**
2.5 Types of Data	**16**
2.6 Editing of Data	**20**

2.7 International Cricket Council (ICC)	**20**
2.8 ICC Player Rankings	**21**
2.9 T-20 Ratings	**21**
2.10 Pre-requisite Tables Related to Present Study	**21**
CHAPTER 3: METHODOLOGY ANDANALYSIS OF DATA	**26**
2.1 Aim of the Project Report	**27**
2.2 Materials used for the Study	**27**
2.3 Approach to the Field	**27**
2.4 Methodology	**27**
2.5 Hypotheses under Consideration	**28**
2.6 Calculation of Mean, Standard Deviation and Coefficient of Variation	**28**
CONCLUSIONS AND SUGGESTIONS	**33**
REFERENCES	**34**

GLOSSARY

Glossary of Cricket Terminology used in this Project

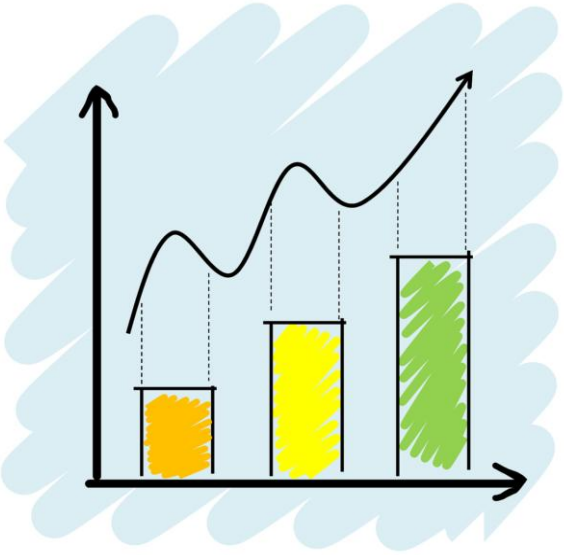
- ✓ **Batsman–** The player who hits the ball and scores runs for the team.
- ✓ **Bowler–** The player who bowls the ball towards the batsman.
- ✓ **Fielder–** The player who fields the ball and either prevents the batsman from scoring runs or tries to catch batsmen out when the ball is hit in the air.
- ✓ **Wicket-keeper–** The player who fields directly behind the batsman and is responsible for catching the ball if it goes past the batsman.
- ✓ **Pitch**– the playing surface on a cricket field where the ball is bowled, and batsmen run.
- ✓ **Over**– A set of six consecutive balls bowled by a single bowler at one end of the pitch. The direction of bowling from one side of the pitch (i.e.: where the wickets are aimed at) alternates, changing after each over is bowled.
- ✓ **Innings–** Innings refers to the session where one team or one batsman bats and scores runs.
- ✓ **Average–**The average runs scored each match by a batsman in his playing career, or the time taken under consideration.
- ✓ **Out**– This is when the bowling team dismisses a batsman in various ways (bowled, caught, run out, leg-before wicket, stumped, timed out or ball tampering). Upon being declared "out", the batsman has to return to his team either in the change room or pavilion.
- ✓ *** = Not Out–** This sign usually denotes cases where a batsman has scored runs but has not got out by the bowling team. It can be used during or after the game to show that the batsman did not go out.
- ✓ **DNB–** Did Not Bat
- ✓ **TDNB–** Did Not Bat
- ✓ **Test matches**– International level of cricket played over five days of play.
- ✓ **One-day International (ODI)–** International level of cricket played over one day of play (50 overs per team).
- ✓ **Twenty20 (T20)**– The shorter version of one-day cricket which has 20 overs per team (innings).

LIST OF FIGURES

FIGURES	PAGE NO.
FIGURE 1.1:A cricket field showing the location of the pitch and some possible fielding positions for the players.	**4**
FIGURE 1.2: The layout of the pitch.	**4**
FIGURE 1.3: The pitch in use, with bowler, batsmen, fielders, and umpire (referee).	**5**
FIGURE 3.1: Graphical Representation of the Average Runs (μ) Scored by the Batsmen.	**31**
FIGURE 3.2: Graphical Representation of the Standard Deviation (σ) of the runs scored by the Batsmen.	**31**
FIGURE 3.3: Comparative Chart of the Mean and Standard Deviations of the runs scored by the respective Batsmen.	**32**

LIST OF TABLES

TABLES	PAGE NO.
TABLE NO. 1: List of Top 20 ICC Batsmen (T-20 Internationals) as on 25th August, 2021.	**22**
TABLE NO. 2: The 10 randomly selected ICC Batsmen from the Top 20.	**23**
TABLE NO. 3: Details of the 10 randomly selected ICC Batsmen from the Top 20.	**23**
TABLE NO. 4: Batsmen and their match-by-match runs.	**25**
TABLE NO. 5: Modified Table of match-by-match runs of the Batsmen.	**31**
TABLE NO. 6: Calculation of Mean(μ), SD(σ) and Covariance (C.V.) of the top 10 T20 cricketers.	**32**



CHAPTER 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1 Sports Analytics

Sports analytics play a major role in various problems associated with sports. Some of these problems are the ranking of individual players and their specialized skills, the composition of teams with an optimal balance of specialized skills, the ranking of teams, the negotiation of contracts, the evaluation of sports businesses and their potential revenue streams, the planning of both physical and mental training, the development of strategies for winning games and tournaments, assessing the effectiveness of coaches and referees, the medical and actuarial aspects of sports injuries (health and insurance), the analysis of existing rules and the need for improving such rules, the improvement of equipment and technology, the determination of awards, the keeping of historical records and the generation of odds for gambling activities. Related to all of the above is the coherent statistical presentation of both raw data and its inferences to the decision makers to facilitate successful planning and implementation. Furthermore, the media and the public have a great appetite for well visualized statistics.

Sports such as baseball, cricket, football, basketball and ice-hockey generate vast sums of money for players, teams, media, advertisers, sponsors, facilities providers and others. The financial imperatives of these activities have given rise to intense statistical analyses. New opportunities for sports analytics have arisen due to the advent and availability of detailed and high-quality data. The technology has been evolving and has led to investigation of problems that had not been previously considered. The data is used in real time for television telecasts but has also been used to analyse the various components of the game. Today's level of sports analytics has evolved where both the technology which provides data, and the statistical methodologies which provide the tools for analysing data, improved very rapidly.

One of the pioneers of modern sports analytics is Bill James who began writing statistically based articles on baseball while working in a factory warehouse. However, many of his articles did not get published because editors felt his style would not have widespread appeal. He then published a series of books titled "Baseball Abstract" beginning in 1977. Although he discontinued writing his self-published books in 1988, he has continued to write books on baseball history. His work was revolutionary because of the use of innovative statistics and his use of play-by-play accounts of games which today is considered absolutely essential for statistical analysis. James

initiated a movement today known as sabermetrics; a word coined from SABR (Society for American Baseball Research). SABR is an organization that boasts over 6500 members, produces baseball research and hosts annual meetings. Sabermetric methods have been used to investigate many problems of interest in baseball including the prediction of the outcomes of matches.

The successful results obtained by the use of statistics and the competitive advantage enjoyed by teams that use these techniques has resulted in a shift towards analytics in major sports. All of the major professional sports associations have statistics on their websites. Due to the onset of sports analytics and the availability of big data there has been a recent increase in analytics research papers. There are also numerous international sports analytics conferences which give a platform to disseminate sports analytics. For example, the annual MIT Sloan Sports Analytics Conference provides a forum for professionals, academics, researchers, students, media and other interested parties to discuss sports analytics and learn about the latest innovations.

Though sports analytics has been rapidly developing, it has not been the case with cricket. Due to historical reasons where cricket was perceived as a leisurely gentleman's game played without remuneration to players (until recently), cricket was not subject to large financial transactions. This has changed in the last few years with the introduction of shorter formats of the game. The shortest and newest format, known as T20, which is explained below, generates intense interest and vast sums of money, especially in the Indian sub-continent. The demand for cricket analytics has increased accordingly and the main website for cricket information and data is cricinfo.com. Since the number of research papers on cricket statistics is limited when compared to more popular sports like basketball, baseball or football,this dearth of research has motivated my own research in this field. This is apart from my avid interest in the game of cricket; as a researcher and as a fan of the sport. Due to this exclusive focus on cricket, I believe that my thesis is unique[1, 12].

1.2 Introductionto Cricket

Cricket is a sport that originated in England in the 16th century and later spread to her colonies. The first international game however did not feature England but was played between Canada and the United States in 1844 at the grounds of the St George's Cricket Club 3 in New York. In time, in both of these countries, cricket took a back seat to other, faster sports like ice-hockey, basketball, and baseball. International cricket is played today by a number of British Commonwealth countries; the main ones being Australia, Bangladesh, England, India, New Zealand, Pakistan, South Africa, Sri Lanka, West Indies

and Zimbabwe. These teams are members of the *International Cricket Council (ICC)*. A second rung of international teams called Associates includes numerous countries including Canada. Cricket is played on an oval-shaped playing field and, apart from baseball, is the only major international sport that does not define an exact size for the playing field. The main action takes place on a rectangular 22-yard area called the pitch in the middle of the large playing field. A diagram showing a cricket field is given in Figure 1.1 and the pitch is magnified in Figure 1.2.

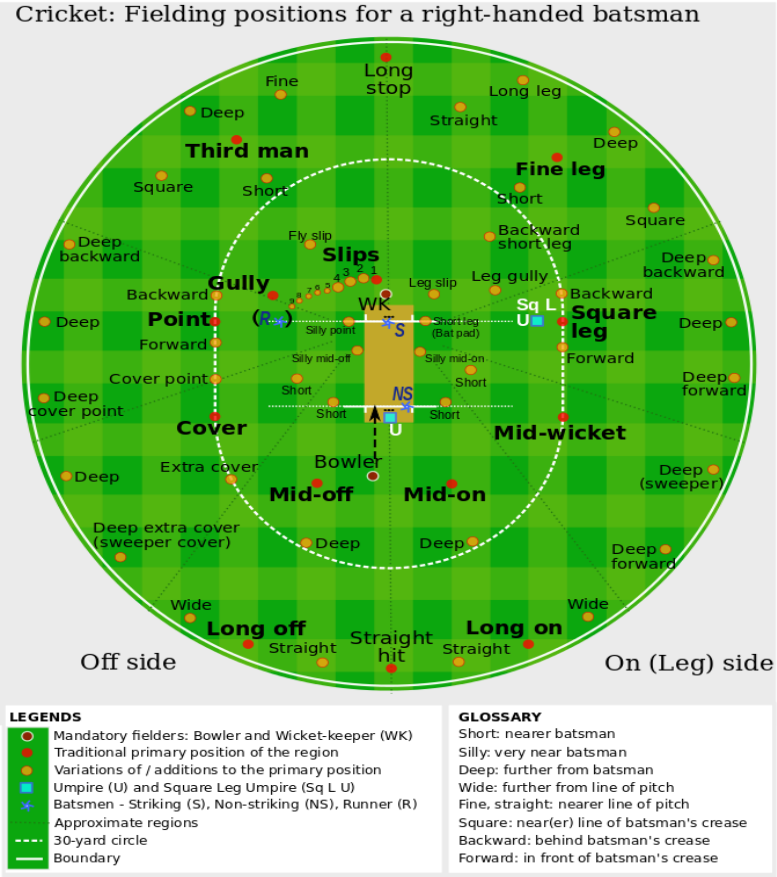


Figure 1.1: A cricket field showing the location of the pitch and some possible fielding positions for players.

Figure 1.2: The layout of the pitch.

Cricket is a game played between two teams of 11 players each, where the two teams alternate scoring (batting) and defending (fielding). A player (bowler) from the fielding team delivers a ball to a player (batsman) from the batting team, who should strike it with a bat in order to score while the rest of the fielding team (fielders) defend the scoring. Furthermore, though it is a team sport, the bowler and batsman in particular, and fielders to some extent, act on their own, each carrying out certain solitary actions independently. The simplicity of these actions (relative to sports such as hockey, soccer and basketball) facilitate statistical modelling. In the process of batting, batsmen can get dismissed (get out) due to a variety of lapses on their part. When all the batsmen from the batting team have been dismissed, or the batting side has faced their allotted number of overs, (each over normally consists of six balls), that team's turn (called innings) is concluded.Their score (number of runs) is recorded.[16] A photo of some game action is provided in Figure 1.3.

Figure 1.3: The pitch in use, with bowler, batsmen, fielders, and umpire (referee).

The teams then change places and the fielding team now gets to wield the bat and try to overtake the score of the team that batted first. At the end of one such set of innings (in the shorter versions of cricket) and two such sets of innings (in the longer version of cricket) the winner is selected on the basis of the most runs scored. This is a very simplified explanation of a very complex game, and there are many variables and constraints that come into play.

When international cricket matured, the standard format was a match that could last up to five whole days. This format is called a *test match*. But even after five days of play the match could end in a draw which means that there is no winner. This was fine in a more leisurely age when both players and

spectators had more time, when playing the game was more important than winning, and when most cricketers were amateur players. But as lifestyles became faster, spectators became ever more reluctant or unable to spend five days watching one match (sometimes with no result). Meanwhile, other faster sports became crowd pullers and earned much more in the way of ticket sales and TV rights. As a result, cricketers became the poor relatives in the sports world. In the 1960s a shorter version of cricket was developed called 'one-day cricket' with each batting side given 65 overs, and later 50 overs in which to score runs.When this format was used in international matches, they became known as *one-day Internationals*or *ODIs*. This version of cricket was much more exciting to watch, as the batsmen had to wield the bat aggressively. Compared to the five-day long test matches, the advent of the 50-over format was a dramatic improvement in terms of spectator entertainment. However, even a 50-over match lasted about 8 hours and could not compete with the two to threehours match times and attention spans of the fans of ice-hockey, football, baseball and basketball. As competition increased for the sports fan's dollar, and TV advertising income was linked directly to the number of viewers, it was inevitable that a shorter format for cricket would emerge. With declining ticket sales and dwindling sponsorships, the *England and Wales Cricket Board (ECB)* discussed the options for a shorter and more entertaining game limited to twenty overs per side and the first official game in this format was played on June 13, 2003. This version became known as Twenty-20. Since then, *Twenty-20* cricket has exploded in popularity with the label "Twenty20" being shortened to T-20. In 2008 the *Indian Premier League (IPL)* was inaugurated using the T-20 format. The participation of dozens of international players in the IPL tournament, and its exciting format, changed the financial aspects of cricket. The *Board of Control for Cricket in India(BCCI)* was already the richest cricket administration in the world but the T20 format brought the IPL almost up to the NBA level in terms of team salaries on a pro-rata basis. In 2014, the IPL brand value was estimated at US$ 3.2 billion. Cricket had finally found its Eldorado in T-20. Nevertheless, test cricket retains a strong following. Many purists claim it is the only authentic version though it is not financially viable.[14, 15]

1.3 Importance of Statistical Data Analysis in Cricket

Cricket is a game of numbers. It is a sport which is not just confined to athletic talent and technical abilities like having good batting and bowling techniques, a good ability to catch and throw the ball, a good hand-eye co-ordination etc. Though strength and strategies do play a very important role in the game but itstill requires much more than that. Even the strategy making process requires constant thinking, plotting and planning based on the already available data and records like match records- the matches won and lost by a particular cricket team, performance of a player in the previous matches- the wickets

taken by a bowler or the runs scored by the batsmen, the number of times a batsman responds in a certain way to a kind of bowling attack, number of stumpings madeetc. Cricket is henceforth, a sport, that generates a variety of statistics.

The capability to dig into cricketing numbers for improvingoverall team performance, maximizing winning chances and studying the business opportunities, overall market, and economics of cricket via data analysis and statistical tools is a big deal. Cricket analytics provides interesting insights into the game and predictive intelligence regarding game outcomes.Real-time data analytics can help in gaining insights even during the game for changing tactics by the team and by associated businesses for economic benefits and growth.

Today, there are rich and almost infinite troves of cricket game records and statistics available, e.g., ESPNcricinfo and cricsheet. These and several such cricket databases have been used for cricket analysis. Media and entertainment platforms along with professional sports bodies associated with the game use technology and analytics for determining key metrics for improving match winning chances:

- batting performance moving average,
- score forecasting,
- gaining insights into fitness and performance of a player against different opposition,
- player contribution to wins and losses for making strategic decisions on team composition etc. [12, 13]

1.4 A Brief Review of Literature

Cricket is, by nature, a multiplex and strategical sport, and is a sport in which statistics feature heavily.The score-book itself is a polished-up notation system. However, it only provides a limited analysis of performance of the players and teams to the coaches, which in turn, can lead to false and/or misleading feedback or decision-making. Therefore, it is important that a notational system identifies a more complete set of technical and tactical indicators required in performance to maximize the feedback to coaches and performers.Many research papers have been published which measured the performance of the players and their predictions. Ahmed et. alhave tested the significance of the consistency of the batsman performances in T20 among the top 10 in T20, ICC ranked cricketer as ofFebruary 2018.[4]Hughes and Bell developed a hand notation system which analysed 20 elite cricketers from England versus Australia ODI series 1997. Within the system, the authors highlighted ball line and length compared to shot selection which resulted in

significant differences found in playing techniques, tactics and styles adopted. [5]Croucher produced 'Cricket-Stat', a combination of both hand and computerised notation which allowed players to identify reasons for success and failure within performance and highlight tactics linked to successful performance. [6]

Duckworth and Lewis developed the D/L method, which is a mathematical formulation designed to calculate the target score for the batting second in a limited overs match interrupted by weather or other circumstances.[10]Barr and Kantor used a new graphical representation together with the strike rate on one axis and the probability of getting out on the other.[2]Barr and Van Den Honert, explained a measure which is based on the average and a consistency. [3] Lemmer showed in his study that the batting average cannot be satisfied in the case of a small number of scores if the player had a large percentage of not out scores. [11]Dey et. al proposed a statistical based multi-criteria decision-making analysis which provides a comparison between the batsmen in three IPL and evaluate the overall performances of batsmen.[7]

1.5 Some Relevant Definitions Used in the Study

1.5.1 Measures of Central Tendency

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. In other words, they give us an idea about the concentration of the values in the central part of the distribution. As such, measures of central tendency are sometimes called measures of central location. They provide us the gist and give a bird's eye view of the huge mass of unwieldly numerical data. Measures of central tendency are very much useful for describing the distribution in concise manner, comparative study of different distributions and computing various other statistical measures such as dispersion, skewness, kurtosis and various other characteristics of a mass of data. The *mean, median* and *mode* are all valid measures of central tendency. [8] Some basic measures of central tendency are:

1.5.1.1 Arithmetic Mean

The mean (or average) is the most popular and well-known measure of central tendency. It can be used with both discrete and continuous data, although its use is most frequently with continuous data. The arithmetic mean of a given set of observations is their sum divided by the number of observations in the data set. So, if X_1, X_2,……,X_n are the given n observations, then their arithmetic mean, usually denoted by $\overline{X}$ is given by:

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

1.5.1.2 Median

The median of a distribution is the value of the variable which divides it into two equal parts. It is the value of the variable such that the number of observations above it is equal to the number of observations below it. Unlike arithmetic mean which is based on all the items of the distribution, the median is only positional average i.e. its value depends on the position occupied by a value in the frequency distribution. The median is less affected by outliers and skewed data. In order to calculate the median, suppose we have the data below:

65 55 89 56 35 14 56 55 87 45 92

We first need to rearrange that data into order of magnitude (smallest first):

14 35 45 55 55 **56** 56 65 87 89 92

Our median mark is the observation in middle. In this case it is 56. It is the middle mark because there are 5 scores before it and 5 scores after it. This works fine when we have an odd number of scores but when you have an even number of scoreswe simply have to take the middle two scores and average the result. So, if we look at the example below:

65 55 89 56 35 14 56 55 87 45

We again rearrange that data into order of magnitude (smallest first):

14 35 45 55 **55** **56** 56 65 87 89 92

Now we have to take the $5^{th}$and 6^{th}score in our data set and average them to get a median of 55.5.

1.5.1.3Mode

Mode is the value which occurs most frequently in a set of observations and around which the other items of the set cluster densely. In other words, mode is the value of the variable which is predominant in the series. In the case of a discrete distribution, mode is the value of x corresponding to maximum frequency. On a histogram it represents the highest bar in a bar chart or histogram.

1.5.2Measures of Dispersion

Measures of central tendency give us an idea of the concentration of the observations about the central part of the distribution. In spite of their great utility, they have their own limitations. If we know the average of a series alone, we cannot form a complete idea about the distribution because it is quite possible that two or more sets of data may have the same average and yet may differ from each other in other respects.Thus, they must be supported and supplemented by some other measures. One such measure is Dispersion.Statistical Dispersion is basically the variability or spread in a variable or a probability distribution. We study dispersion to have an idea of the homogeneity (compactness) and heterogeneity (scatter) of the distribution i.e. the variation or scatteredness of individual values from their central values such as mean, median and mode. The methods which are used to measure this variation in numerical term are called measures of dispersion or measures of variability. A measure of statistical dispersion is a real number that is zero (0); if all the data are identical, and increases as the data becomes more diverse. It cannot be less than zero. There are two types of measure of dispersion such as absolute measures of dispersion and relative measure of dispersion.[8]

1.5.2.1Absolute Measures of Dispersion

The measures of dispersion which are expressed in terms of the original units of a series are termed as Absolute Measures. These measures give us an idea about the amount of dispersion in a set of observations i.e. express the spread of observations. They give the answers in the same units as the units of the original observations. When the observations are in kilometres, the absolute measure is also in kilometres. Absolute dispersion cannot be used to compare the variation in two or more than two sets of data.The absolute measures of dispersion are classified into two broad categories:

1) Also known as distance measures, these measures are those which express the spread of observations in terms of distance between the values of selected observations., e.g., *Range* and *Quartile Deviation* and

2) the measures which express the spread of observations in terms of average of deviations of observations from some central value, e.g., *mean deviation* and *standard deviation*.

1.5.2.2Relative Measure of Dispersion

To compare the variability of the two series which differ widely in their averages or which are measured in different units, we calculate the

coefficients of dispersion or the relative measures of dispersion. These measures are a pure number in the form of a coefficient and are calculated for the comparison of dispersion in two or more than two sets of observations. These measures are obtained as ratios or percentagesand hence are independent of the units in which the original data is measured. Each absolute measure of dispersion can be converted into its relative measure. Thus, the relative measures of dispersion are:

- *Coefficient of Range.*
- *Coefficient of Quartile Deviation.*
- *Coefficient of Mean Deviation.*
- *Coefficient of Standard Deviation.*
- *Coefficient of Variation.*

1.5.3 Scales for the Level of Measurement

The measures of central tendency and measures of dispersion can be used to describe the values of cases on every variable. The choices depend on the variable's level of measurement. Modes and percent distributions are relatively simple in nature and only require that the attributes that make up a variable be different. This is a property that the attributes of nominal, ordinal, and interval/ratio variables all have. Therefore, we can calculate modes and percentage distributions for all three levels of measurement. Medians, minimums, maximums, ranges, and percentiles provide more information about the scores on variables which only make sense if the attributes of a variable have rank order which is a property possessed by the attributes of ordinal and interval/ratio variables. Thus, we may calculate medians, minimums, maximums, ranges, percentiles, and inter-quartile ranges for ordinal and interval/ratio variables. Means, variances, and SD provide still more information about the scores on variables, but these statistics require the attributes of the variable to form a numeric scale with a fixed unit of measurement. Since only interval/ratio variables have this property, means, variances, and standard deviations may only be calculated for interval/ratio variables.

1.6 Problem and Purpose of the Project

The problem and purpose of the study is to translate a real-life problem into a statistical problem and then apply some suitable statistical techniques to draw inferences. In this case study, wedetermine the most consistent of the10 randomly selectedfrom the top 20 ICC ranked batsmen based on their performances of the past three years i.e., 25thAugust 2018 to 25th August 2021 in T-20, as on 25thAugust 2021.The thing that we admire in our cricketers is their consistency. In cricket, the performances of the batsmen are determined using their averages (Arithmetic Mean) of the runs they have scored over their careers.

However, here, we specifically focus the study on the recordsof the batsmen's match by match run scores of the past three years because the ICC ranking of the players are determined on basis of the present/recent form of the players. Also, the mean of a set of data alone is not sufficient to determine the consistency.Full marks to the gritty player who scores 50 on a minefield, even though he gets out for 50 when well set on a featherbed. But do we admire so much his team-mate who gets a duck in the first instance, but makes amends by crashing an impressive 100 in the second? They have the same average - but do they provide the same value?So, to figure this out, the obvious next thing to do is calculate the standard deviation which will describe how 'spread out' the batsmen's scores are.

Consistency can bemeasured by calculating the standard deviation, which, in simple terms, seeks to measure the average deviation that each score is from the overall mean. The lower the standard deviation, the lower the variation in the scores. In other words, better batsmen (i.e., those with higher averages) have more 'space' to spread into, which will result in larger standard deviations. On this measure alone, the 'most consistent' performers are likely to be the ones with the lowest average scores. To meet the stated objective, secondary data has been collected from the relevant sources and the mean, variance and coefficient of variation has been calculated.

1.7 Chapter-wise Summary

1.7.1 CHAPTER 1: Introduction

Chapter 1 is an introduction to the proposed study of the project work and gives a brief overview of the problem and purpose of the project. It contains some relevant definitions and concepts used in the successful completion of the project.

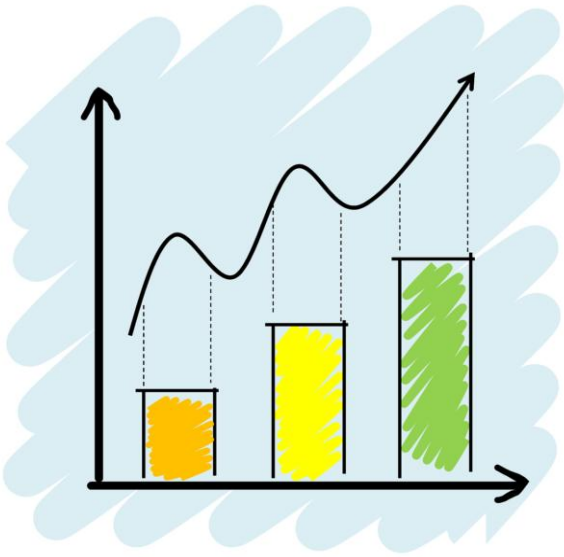
1.7.2 CHAPTER 2: Data Collection

Chapter 2 focusses on the various methods of data collection for the project and organisation of the data in a more convenient way. It also contains additional information on the subject of study and the tabulation of the data obtained from various sources.

1.7.3 CHAPTER 3: Methodology and Analysis of Data

Chapter 3 contains the methodology opted in the analysis of the data and the analysis of the data to obtain the required results. Here, the mean, variance and

coefficient of variation of the 10 randomly selected batsmen are determined on the basis of which the conclusions will be drawn.



CHAPTER 2

DATA COLLECTION

CHAPTER 2

DATA COLLECTION

2.1Research Methodology

Research methodology is a way to systematically solve the research problem. It is the set of specific procedures or techniques used to analyse information about a subject under study. It may be understood as the science of studying how a research is conducted scientifically. The steps to solve a research problem are listed below:

- First, the variable is identified andthen determined whether that it is quantitative or qualitative.
- Then the summary measure of interest is identified which could be a mean, a proportion, a difference in means, a difference in proportions, a regression coefficient etc. The summary measure would partially depend on the measure is qualitative or quantitative.
- Next, the null hypothesis is set up and an alternative hypothesis that would accept in case the null hypothesis is rejected. Decision is made whether thealternative hypothesis is one-sided or two-sided.
- The statistical distribution of the chosen summary measure for the sample in hand is assessed. If the sample size is large, many of such distributions would be approximately Gaussian. If the sample size is sample, a non-parametric test or exact procedures may be needed.
- The level of significance, 5% or whatever is considered appropriate, is fixed.
- The appropriate statistical test criterion based on the above-mentioned considerations are identified.
- If necessary, the power of the test for the specified difference are also calculated.

2.2Objectives of the Present Study

Solving statistical problems with compiled data, using any means of calculation is generally what we do. But the difficulty arises when we are asked to solve problems based on real life instances. To do that, we need to first translate the problem in to a statistical problem i.e., we have to formulate the problem on our own and only then we can apply

some suitable statistical techniques and draw inferences.So, the general objectives to be attained in the present study are as follows:

- To learn how to formulate a real-life problem.
- To learn how to translate the problem to a statistical problem.
- To learn how to collect data, though secondary.
- To be familiar with handling large data using software.
- Analysing the data using statistical methods.
- Learning how to interpret the findings.

2.3Data Collection

The process of counting or enumeration or measurement together with the systematic recording of results is called the collection of statistical data. The entire structure of a statistical analysis for any enquiry is based upon the systematic collection of data. Hence, the first step in any enquiry (investigation) is the collection of data. The well-trained person who conducts the statistical enquiry is known as an investigator or an enumerator. The respondents are the persons from whom the information is collected.

The purpose of data collection is to obtain information to keep on record, to make decisions about important issues, or to pass information on to others. Prior to any data collection, pre-collection activity is one of the most crucial steps in the process. It is often discovered too late that the value of their interview information is discounted as a consequence of poor sampling of both questions and informants and poor elicitation techniques. After pre-collection activity is fully completed, data collection in the field, whether by interviewing or other methods, can be carried out.

Other main types of collection include census, sample survey, and administrative by-product. In the census method, we resort to 100% inspection of the population and enumerate each and every unit of the population and has advantagessuch as accuracy and detail and disadvantages such as cost and time. A sampling is a data collection method that includes inspecting only aselected representative of the population and has advantages such as cost and time and disadvantages in terms of accuracy and detail. A choice between the sample method and the census method is to be made depending upon the objectives and scope of the survey, the limitations of resources in terms of time, money, manpower, etc., and the degree of accuracy desired.[9]

2.4Collection of Statistical Data

A sequence of observation, made on a set of objects included in the sample drawn from population is known as statistical data. Statistical data are numerically expressed aggregate of facts, collected in a systematic manner for a predetermined purpose and placed in related to each other. Statistical data are either enumerated or estimated.

- ***Ungrouped Data*:** Data which have been arranged in a systematic order are called raw data or ungrouped data.
- ***Grouped Data*:** Data presented in the form of frequency distribution is called grouped data.

2.5Types of Data

The proper and correct collection of data is the most important factor in any statistical enquiry. Statistical data are of two types:

(i) Primary Data

(ii) Secondary Data[9]

2.5.1 Primary Data

The primary data are the data originally collected by an investigator or agency and used by them for a specific purpose. In other words, primary data are those data which are collected for the first time and thus are original in character because they have never been used for any other purpose earlieror have undergone any sort of statistical treatment. For example,Reserve Bank of India Bulletin (monthly), annual report of railway board etc.

2.5.1.1 Methods of Collecting Primary Data

Primary data are collected by the following methods:

- ***Direct Personal Investigation:*** This method consists in the collection of data personally by the investigator (or organising agency) from the sources concerned. The investigator establishes personal contact with the informants or respondents and conducts on the spot enquiry. The data collected this way is usually accurate and reliable because of personal approach. This method of collecting data is suitable when the field of investigation is not too wide in terms of

the number of persons to be interviewed or the area to be covered.

- ***Indirect Oral Investigation:***This method is used in cases where 'direct personal investigation' is not practicable because the informants are reluctant to furnish the requisite information or due to the extensive nature of the enquiry or due to the fact that direct sources of information do not exist or are unreliable. Under this method, indirect evidence of third parties who are in touch with the desired facts, is record. This method is economic, time saving and a wide area can be brought under investigation as compared to the 'direct personal investigation' method.
- ***InformationReceived Through Local Agencies:***There is no formal collection of any data but the local correspondents residing in different areas collect the information and report the same to the authority. It is method usually employed by newspaper or periodical agencies who require information in different fields like sports, riots, strikes, accidents, economic trends, policies, business stocks, share market etc. Though this method works out to be very cheap but gives only rough estimate of the requisite information.
- ***Mailed Questionnaire Method:***The schedules of questions called questionnaires are mailed to the respondents or informants with the request of quick response within the specified time after duly filled in. This method is usually adopted by research institutes and private bodies. This method is the most economical in terms of time, money and manpower. The bias of the investigator is completely ruled out and original data are collected.
- ***Schedules sent through Enumerators:*** Schedule is the method of getting answers to the questions in a form which are filled by the interviewers in a face-to-face situation. In this method, the enumerators go to the respondents with the schedule and record their responses.

Population census, all over the world is conducted by this technique. This method is helpful in obtaining maximum possible responses but is fairly expensive and time consuming as compared to the other methods.

2.5.2 Secondary Data

The secondary data are the information which have been previously collected and processed by some agency or person for their own purpose but are now used in a different connection. In other words, the primary data used and analysed again to extract some other information are termed as secondary data. So, after statistical treatment, the primary data lose their original shape and become secondary data. Primary data once published becomes secondary data.The secondary data are not pure in character, i.e., they are like finished products since they have undergone some treatment at least once.For example,Monthly Abstract of Statistics, International Labour Bulletin(monthly) etc.

2.5.2.1 Methods of Collecting Secondary Data

The secondary data are collected using the following sources:

- ***Official Publications:*** The various government organisations collect, compile and publish statistical data on a number of topics of current interest- prices, wages, population, labour, trade, army etc. Some of the examples of such government organisations are: Office of the Registrar General and Census Commissioner of India, New Delhi(1949), Labour Bureau- Ministry of Labour(1946), Directorate of Economics and Statistics- Ministry of Agriculture and Irrigation(1948) etc.
- ***Semi-Official Publications*:** Information from the publications of the semi-government statistical organisations are very useful. Some of them are: Statistical department of Reserve Bank of India (Mumbai) which brings out an Annual Report of the Bank, Currency and Finance,

The Institute of Economic Growth, Delhi etc.

- ***Publications of Research Institutions*:**Individual research scholars, the different departments in the various universities and various research organisations publish the findings of their research programmes in the form of research papers or monographs or journals . For example- Indian Statistical Institute (I.S.I), Delhi and Kolkata, National Council of Educational Research and Training (N.C.E.R.T) etc
- ***Publications of Financial and Commercial Institutions:*** Many private commercial and trade institutions publish reports and statistical material on the current economic, business and other phenomena. For example- Institutes of Chartered Accountants of India, bank bodies etc.
- ***Reports of various Committees and Commissions appointed by the Government:*** The report of the surveys conducted by the committees and commissions set up by the government to take in consideration the views of the experts on some important matters like wages, land, education, prices, Dearness Allowances etc. For example- Pay Commissions Reports, Simon-Kuznets Committee report on National Income in India etc.
- ***Newspapers and Periodicals:***Statistical information and materials on a number of current social-economic problems can be obtained from the numerical data published in the newspapers, journals or magazines. For example- The Financial Express, Economic Times, etc.
- ***International Publications:*** The publications of a number of foreign governments or international agencies provide crucial statistical data on very important current topics. For example-

World Bank, World Health Organisation (W.H.O) on contagious diseases etc.

- ***Unpublished Sources:*** Not always the statistical data are published. Private firms, business enterprises, various departments of the government, researches carried out by individual research scholars in their universities or research institutes who may not like to release their data to any outside agency are the sources to a number of unpublished statistical materials and records.

2.6Editing of Data

After the collection of data using either the primary sources or secondary sources, the next step is scrutinising and editing of the data. Editing means the examination of collected data to discover any error and mistake before presenting it for statistical analysis. Editing of the data checks for completeness of data, consistency of data, accuracy of data and homogeneity of data. The desired degree of accuracy and the extent to which the errors can be tolerated in the enquiry have to be pre-decided. The editing of secondary data is simpler than that of primary data.

2.7International Cricket Council (ICC)

The International Cricket Council (ICC) is the globalgoverning body of cricket. It was founded as the Imperial Cricket Conference in 1909 by representatives from England, Australia and South Africa, later renamed as the International Cricket Conference in 1965and took up its current name in 1989.

Representing 106 members, the ICC governs and administrates the game and works with the members to grow the sportby creating more opportunities for more people and nations to enjoy it and increase the competitiveness of international cricket at all levels. The ICC is also responsible for the staging of all ICC Events, organising and governing the major international cricket tournaments, most notably the Cricket World Cup.It also appoints the umpires, referees and all match officials that officiate at all sanctioned Test matches, One Day International (ODI) and Twenty20 Internationals (T-20).

It presides over the ICC Code of Conduct, playing conditions, the Decision Review System and other ICC regulations. Through its Anti-Corruption and Security Unit (ACSU), it coordinates action against corruption and match fixing.The ICC does not control bilateral fixtures

between member countries, it does not govern domestic cricket in member countries, and it does not make the laws of the game, which remain under the control of the Marylebone Cricket Club.

2.8ICC Player Rankings

The ICC Player Ratings are a sophisticated moving average. Players are rated on a scale of 0 to 1000 points. The points/ratings of a player increase if his performance is improving based on his past records.Again, if his performance is deteriorating, then his points will also decrease accordingly. Ranks are assigned to the players based on these ratings. The value of each player's performance within a match is calculated using an algorithm, a series of calculations (all pre-programmed) based on various circumstances in the match. Players have to have appeared in a match within the qualifying period to appear in the lists (normally 12-15 months for Tests, 9-12 months for T-20s and ODIs). All of the calculations are carried out using pre-programmed formulae, using the information published in a Test match scorecard. There is no human intervention in this calculation processand no subjective assessment is made of the quality of the pitch or of the players.[18, 19]

2.9T-20 (Twenty-20) Ratings

ICC has set up a criterion to rank the players according to their performances. According to ICC, teams will need to play six matches against other teams in the previous three to four years to remain in the rankings table.After the T-20 match, the two teams involved receive points based on a mathematical formula. The total of each team's points total is divided by the total number of matches to give a rating, and all the teams are ranked on a table in order of rating. Each team scores points based on the results of their matches over the last 3−4 years − all matches played in the 12–24 months since the May before last, plus all the matches played in the 24 months before that, for which the matches played and points earned both count half. Each May, the matches and points earned between 3 and 4 years ago are removed, and the matches and points earned between 1 and 2 years ago switch from 100% weighting to 50% weighting.[18, 19]

2.10 Pre-requisite Tables Related to Present Study:

The names,ICC rankings and the ratings of top 20batsmen for T-20 Internationals are shown in the following table:

Table No. 1: List of Top 20 ICC Batsmen (T-20 Internationals) as on 25th August, 2021

Rank	Name of the Players	Ratings
1	Dawid Malan	841
2	Babar Azam	819
3	Devon Conway	774
4	Aaron Finch	733
5	Virat Kohli	717
6	Kannur Lokesh Rahul	699
7	Mohammad Rizwan	692
8	Martin James Guptill	688
9	Evin Lewis	664
10	Rassie van der Dussen	650
11	Hazratullah Zazai	636
12	Quinton de Kock	609
13	Eoin Morgan	599
14	Jason Roy	592
15	Colin Munro	592
16	Jos Buttler	590
17	Jonny Bairstow	588
18	Henry George Munsey	583
19	Rohit Sharma	577
20	Rahmanullah Gurbaz	576

The details and match-by-match runs of the 10 randomly selected batsmen out of the top 20 for T-20I are shown in the following tables:

Table No. 2: The 10 randomly selected ICC Batsmen from the Top 20

Sl. No.	Randomly Selected Ranks	Name	Playing Nation	Ratings
1	8	Martin James Guptill	England	688
2	18	Henry George Munsey	Scotland	583
3	10	Rassie van der Dussen	South Africa	650
4	1	Dawid Malan	England	841
5	5	Virat Kohli	India	717
6	4	Aaron Finch	Australia	733
7	16	Jos Buttler	England	590
8	9	Evin Lewis	West Indies	664
9	7	Mohammad Rizwan	Pakistan	692
10	19	Rohit Sharma	India	577

Table No. 3: Details ofthe 10 randomly selected ICC Batsmen from the Top 20

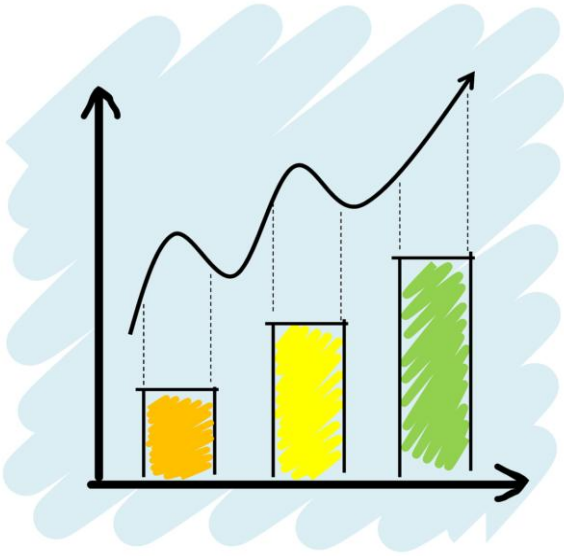
Sl. No.	Full Name	Playing Nation	Born	Playing Role	Batting Style	Bowling Style
1	Martin James Guptill	New Zealand	30.09.1986 Auckland, New Zealand	Opening Batsman	Right-handed	Right arm offbreak
2	Henry George	Scotland	21.02.1993 Oxford,	Opening Batsman	Left-handed	Right-arm

	Munsey		England			medium-fast
3	Rassie van der Dussen	South Africa	07.02.1989 Pretoria, Transvaal, South Africa	Top-order Batsman	Right-handed	Right-arm legbreak
4	Dawid Malan	England	03.09.1987 Roehampton, England	Top-order Batsman	Left-handed	Right-arm legbreak
5	Virat Kohli	India	05.11.1988 Delhi, India	Top-order Batsman	Right-handed	Right arm medium
6	Aaron Finch	Australia	17.11.1986 Colac, Victoria, Australia	Top-order Batsman	Right-handed	Slow left arm orthodox
7	Jos Buttler	England	08.09.1990 Taunton, Somerset	Wicketkeeper Batsman	Right-handed	----------
8	Evin Lewis	West Indies	27.12.1991 Port of Spain, Trinidad	Opening Batsman	Left-handed	----------
9	Mohammad Rizwan	Pakistan	01.06.1992 Peshawar, Khyber Pakhtunkhwa, Pakistan	Wicketkeeper Batsman	Right-handed	----------
10	Rohit Sharma	India	30.04.1987 Bansod, Nagpur, Maharashtra, India	Top-order Batsman	Right-handed	Right arm offbreak

Table No. 4: Batsmen and their match-by-match runs

Matches	M.J. Guptill (1)	H.G. Munsey (2)	R. van der Dussen (3)	D. Malan (4)	V. Kohli (5)	A. Finch (6)	J. Buttler (7)	E. Lewis (8)	M. Rizwan (9)	R. Sharma (10)
1	1	32	56	11	4	1	13	18	2	6
2	11	50	13	39	TDNB	0	15	1	1*	111*
3	DNB	16	0	55	61*	3	2	89	26	4
4	2	127*	45	103*	24	1	57	0	31	7
5	41	34	41	11	72*	7	44	0	14	TDNB
6	33	27	34	23	19	27	77*	10	0	23
7	27	6	64	54*	28	0	7	68	5*	1
8	50	46	1	7	59	28	22	14	DNB	50
9	30	23	DNB	66	72*	0	67*	24	TDNB	38
10	33	14	31	42	9	8	28	40	DNB	5
11	31	11	43*	21	94*	64	0	40	0*	24
12	4	51	11	19	19	0	83*	DNB	DNB	67
13	2	7	6	55	70*	37	9	53	DNB	12
14	5	65	37	99*	TDNB	37*	52	2*	17	9
15	34	17	24	24*	30*	17	68*	46	22	9
16	TDNB		37	24	26	52*	59	28	89	85
17	6		25*	18	45	42	21	6	104*	2
18	21		74*	14	11	14		21	51	8
19	19		34*	68	38	55		71	42	15
20	0		52	7	11	46		21	74*	71
21	97		56*	4	9	40		27	0	7
22	43		2	76	40	39		7	73*	8
23	7		32	1	85	35		52	0	65
24	71		6	1	0	0		0	82*	60*
25	35		1	31	73*	1		31	13	15
26	21		25		77*	12		79	91*	12
27	44		6		1	69		6	63	64
28					80*	79*		35*	37	
29						36			76*	
30						4			TDNB	
31						6			46	
32						30			TDNB	
33						53			TDNB	
34						34				

Total	**668**	**526**	**756**	**873**	**1057**	**877**	**624**	**789**	**959**	**778**



CHAPTER 3

METHODOLOGY AND ANALYSIS OF DATA

CHAPTER 3

METHODOLOGY ANDANALYSIS OF DATA

3.1 Objective of the Project Report

The basic purpose of this research study is to analyse the consistency of the 10 randomly selected out of the top 20 ICC ranked batsmenand determine the most consistent of the 10 for the last 3 years i.e., 25thAugust 2018 to 25th August 2021 using Measures of Central Tendency and Measures of Dispersion.

3.2 Materials used for the Study

To accomplish the objective of analysing the consistency of the 10 randomly selected out of the top 20 ICC ranked batsmen,secondary data regarding the ICC match-by-match runs for the last three years of the batsmen has been collected from the relevant websites as of August, 2021.[17]

3.3 Approach to the Field

As a curriculum for the B.A./B.Sc. of the students in TDC 6th Semester having Major in Statistics, the departmental management committee generally decides the topics and accordingly the place for data collection. I have prepared my project report under the supervision of Dr. N. Ahmed. And the secondary data has been collected from the website www.stats.cricinfo.com.

3.4 Methodology

In this project report, an attempt has been made to study the consistency of the randomly selected 10 out of the top 20 ICC ranked batsman in T20 (Twenty-20) for the last three years i.e., 25th August 2018 to 25th August 2021. To meet the stated objective,the secondary data has been collected from the verified sources. First, the top 20 ICC ranked batsmen (T-20I) has been recorded along with their ranks and ratings. Using the random number table, 10 out of the top 20 batsmen are randomly selected for the analysis. This random selection of the 10 players using random number table gives a Simple Random Sample and the method is known as Simple Random Sampling (SRS).The

details of the 10 randomly selected batsmen are recorded. The statistical analysis has been carried out by using simple descriptive statistics such as measures of central tendency and measures of dispersion.

For the calculation of the average runs scored by a batsman, the number of times ‘notout,‘DNB’ and ‘TDNB’ are omitted from the number of innings played i.e.,

Adjusted Number of Innings Played (n) = Original Number of InningsPlayed−No of times ‘not out’−no of times ‘DNB’ and ‘TDNB’

Also, a modified table of the batsmen’s match-by-match runs are preparedwhere the ‘DNB’ and ‘TDNB’ matches are ignored. By replacing the ‘DNB’s and ‘TDNB’s with zero, the results in the computation of standard deviation are affected and hence it was decided to ignore the innings with DNB and TDNB as a whole. In the modified table, theruns scored in the innings where the batsmenwere ‘notout’are added to the runs scored in the preceding innings or the succeeding innings to obtain a more accurate measure of the mean and standard deviation. Also,a player who has scored an average run of more than 30 is considered to be a good player. The mean, standard deviation and graphical representation were considered indrawing conclusions about the performances of the batsmen.

3.5 Hypotheses Under Consideration

Below stated are the hypotheses in consideration:

- A player who scored an average run of more than 30 is considered to be a good player.
- A player who has minimum dispersion i.e., tends to zero is considered to be a reliable and consistent player.

3.6 Calculation of Mean, Standard Deviation (SD) and Coefficient of variation (CV):

The analysis is done purely on the basis of the Table No. 4 defined in the Chapter 2. A modified table of the runs scored by the batsmen is made by ignoring the DNB and TDNB. Also, the runs scored in the innings where the batsmen were not-out is added to runs scored in the preceding innings or the succeeding innings (in this report, the preceding innings).Also, the Mean, SD and CV are calculated usingthe following formulas respectively:

Mean (μ or $\overline{X}$),

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Standard Deviation (σ),

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (X - \overline{X})^2}{n}}$$

Coefficient of Variation,

$$CV = \frac{\sigma}{\mu}$$

The modified table is as follows:

Table No. 5: Modified Table of match-by-match runs of the Batsmen

Matches (n)	**M.J. Guptill (1)**	**H.G. Munsey (2)**	**R. van der Dussen (3)**	**D. Malan (4)**	**V. Kohli (5)**	**A. Finch (6)**	**J. Buttler (7)**	**E. Lewis (8)**	**M. Rizwan (9)**	**R. Sharma (10)**
1	1	32	56	11	65	1	13	18	3	117
2	11	50	13	39	96	0	15	1	26	4
3	2	143	0	158	19	3	2	89	31	7
4	41	34	45	11	28	1	57	0	14	23
5	33	27	41	77	131	7	121	0	5	1
6	27	6	34	7	103	27	7	10	17	50
7	50	46	64	66	119	0	89	68	22	38
8	30	23	1	42	26	28	28	14	193	5
9	33	14	74	21	45	0	83	24	51	24
10	31	11	11	19	11	8	9	40	116	67
11	4	51	6	178	38	64	120	40	73	12
12	2	7	37	24	11	0	59	55	82	9
13	5	65	24	18	9	74	21	46	104	9
14	34	17	170	14	40	69		28	63	85
15	6		108	68	85	42		6	113	2
16	21		2	7	0	14		21	46	8
17	19		32	4	150	55		71		15

18	0		6	76	81	46		21		71
19	97		1	1		40		27		7
20	43		25	1		39		7		8
21	7		6	31		35		52		125
22	71					0		0		15
23	35					1		31		12
24	21					12		79		64
25	44					148		41		
26						36				
27						4				
28						6				
29						30				
30						53				
31						34				
Total	**668**	**526**	**756**	**873**	**1057**	**877**	**624**	**789**	**959**	**778**

Now the table below shows the average runs, standard deviation and coefficient of variation of each individual batsman:

Table No. 6: Calculation of Mean(μ), SD(σ) and Covariance (C.V.) of the top 10 T20 cricketers

Sl. No.	Players	Total Runs	Adjusted Innings (n)	Average (μ)	$\sum(X-\mu)^2$	S.D. (σ)	C.V. ($\frac{\sigma}{\mu}$)
1	**M.J. Guptill**	668	25	26.72	13215.04	22.99	0.86
2	**H.G. Munsey**	526	14	37.57	16257.42	34.08	0.90
3	**R. van der Dussen**	756	21	36	34916	40.78	1.13
4	**D. Malan**	873	21	41.57	47543.14	47.58	1.14
5	**V. Kohli**	1057	18	58.73	36801.61	45.21	0.77
6	**A. Finch**	877	31	28.29	30868.38	31.55	1.11
7	**J. Buttler**	624	13	48	22382	41.49	0.86
8	**E. Lewis**	789	25	31.56	16290.16	25.52	0.80
9	**M. Rizwan**	959	16	59.93	40188.93	50.11	0.83
10	**R. Sharma**	778	24	32.41	31165.83	36.03	1.11

The average runs and the standard deviation of the runs scored by the individual players can be represented graphically as follows:

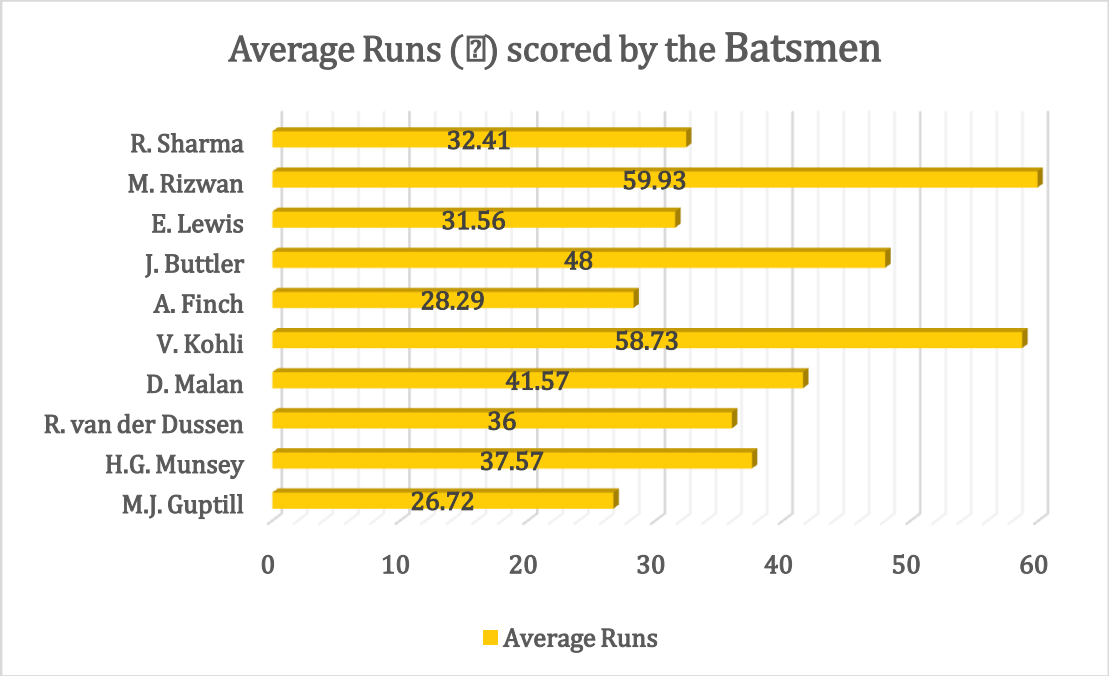


Figure 3.1:Graphical Representation of the Average Runs (μ) Scored by the Batsmen

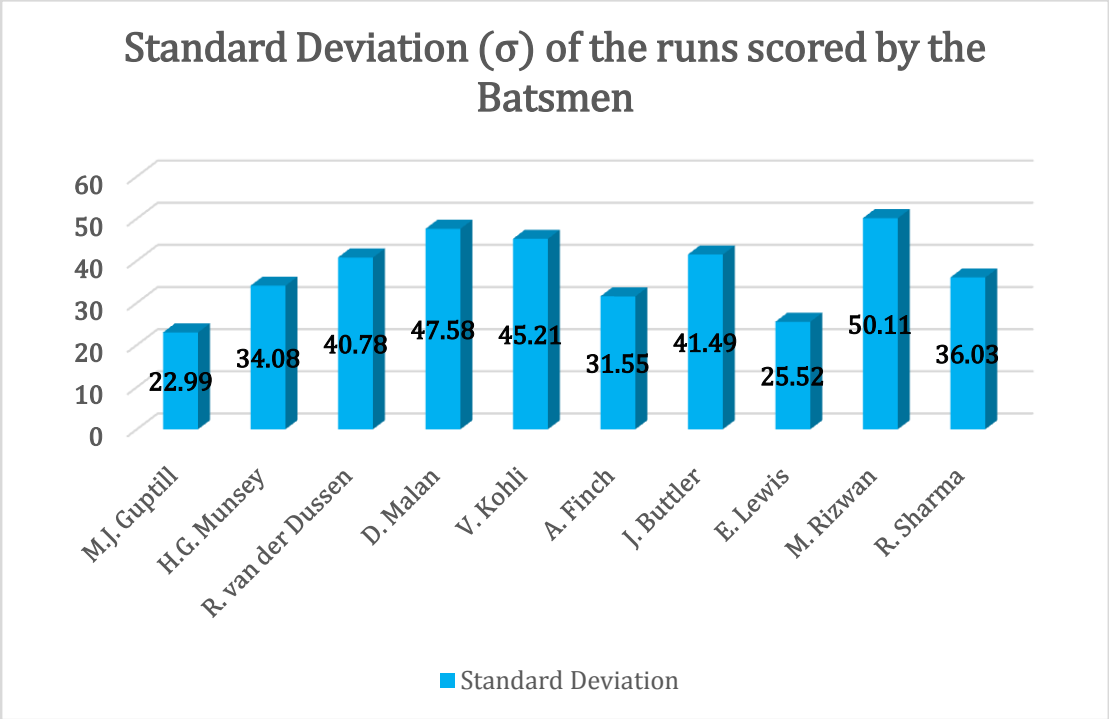


Figure 3.2: Graphical Representation of the Standard Deviation (σ) of the runs scored by the Batsmen.

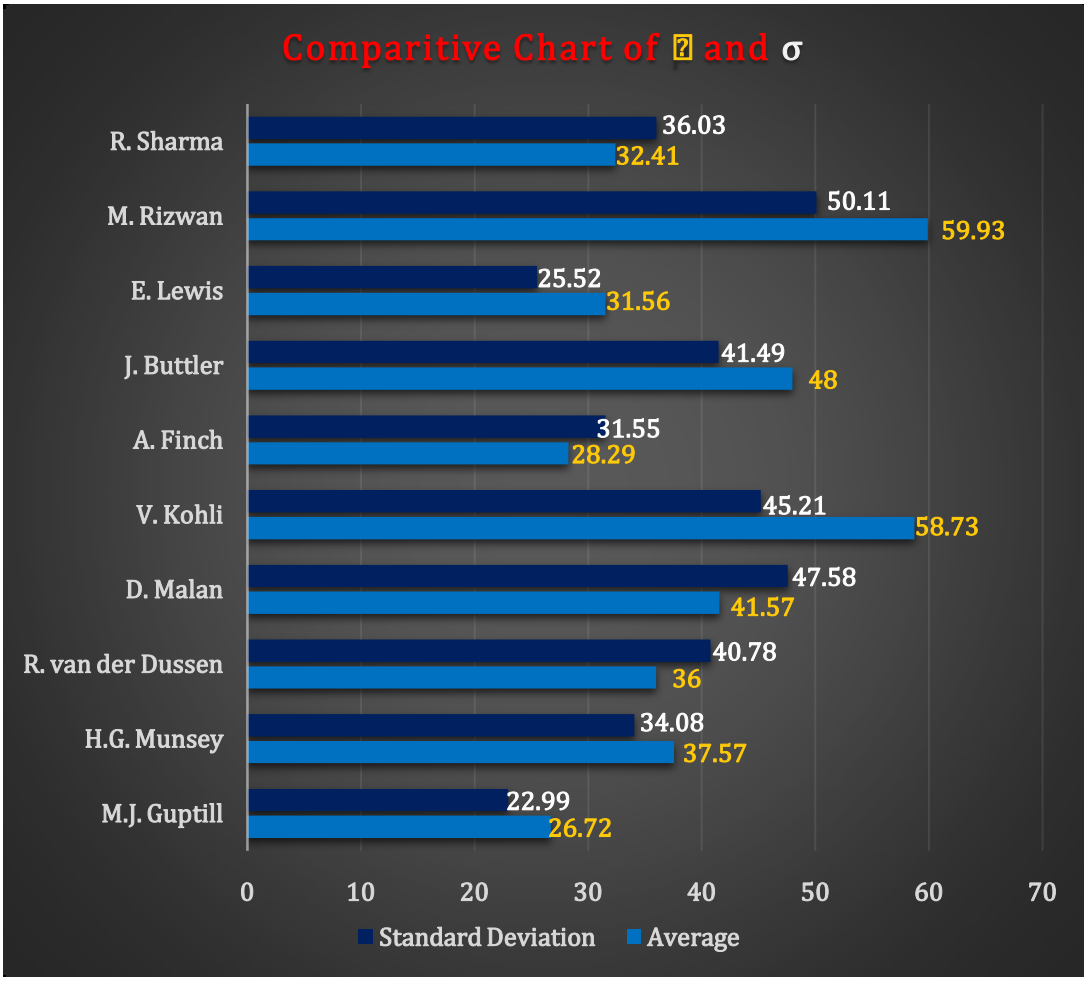


Figure 3.3: Comparative Chart of the Mean and Standard Deviations of the runs scored by the respective Batsmen.

CONCLUSIONS AND SUGGESTIONS

Conclusion:

According to the hypotheses under consideration, we conclude that eight out of ten are good batsmen since their averages are more than 30. The eight batsmen are as follows: Henry George Munsey, Rassie van der Dussen, Dawid Malan, Virat Kohli, Jos Buttler, Evin Lewis, Mohammad Rizwan and Rohit Sharma. The other two players Martin James Guptill and Aaron Finch have averages lower than 30 and hence cannot be declared as good players based on their performances in the last three years i.e., 25th August 2018 to 25th August 2021.

Now comparing the standard deviations of all the ten players, we observe that the standard deviation of Martin James Guptill (σ = 22.99) playing for New Zealand is the lowest as compared to the other batsmen and therefore we conclude that Martin James Guptill is the most consistent or uniformperformer among the randomly selected top ICC ranked T-20 batsmen as on August, 2021. Also, we may expect from Guptill that he can score his average runs (i.e., 26.72) more often.

Next followed by Evin Lewis and Aaron Finch of West Indies&Australia are the next most consistent players with the standard deviations 25.52 and 31.55 respectively. The two Indian Batsmen Rohit Sharma and Virat Kohli are at 5th and 8th position respectively.

Suggestions:

The Standard Deviation (SD) can be shown as a career statistic of a batsmanon the television screens so that one can measure his/her dependability and consistency as a player over another player based on their standard deviations. But as a disadvantage, if it shown on the TV screen, then a gambler can assess andbet more money on those players.

REFERENCES

[1] J. Davis, H. Perera, and T.B. Swartz. Player evaluation in twenty20 cricket. Journal of Sports Analytics, 2015.

[2] Barr, G.D.I. and Kantor, B.S. "A criterion for comparing and selecting batsmen in limited overs cricket". *Journalof the Operational Research Society*. 55 (2004), 1266-1274, Print.

[3] Barr, G.D.I. and Van Den Honert, "R. Evaluating batsmen's scores in test cricket". *South African Statistical Journal*. 32 (1998), 169-183.

[4] Ahmed, N., Sen, P., & Bhattacharjee, N. (2018). ANOVA for Testing the Consistency amongst the Top Ten T20 ICC Ranked Batsmen, 6(1), 48–52.

[5] Hughes, M. & Bell, K. (1998). Performance profiling in cricket, cited in Notational Analysis of Sport (edited by M. Hughes and F. Tavares), pp. 176-183. Porto: Faculty of Sport Sciences, University of Porto, Portugal.

[6] Croucher, J.S. (1987). A guide to cricket-stat: A system for the statistical analysis of cricket matches. Unpublished paper, cited in Yates, C. (2003), A comparison of pattern of play between two international Cricket teams and two playing conditions in one day cricket. Unpublished BSc dissertation. Cardiff, UK: U.W.I.C. 67 pages.

[7] Dey, P.K., Ghosh, D.N., Mondal, A.C. (2012): Statistical Based Multi-Criteria Decision-Making Analysis for Performance Measurement of Batsmen in Indian Premier League, International Journal of Advanced Research in Computer Science, Volume 3, No. 4, 51 - 57.

[8] Gupta, S.C., Kapoor, V.K. (2002): *Fundamentals of Mathematical Statistics* (11thRevised ed., pp. 2.1-2.74), Sultan Chand & Sons.

[9] Gupta S.C. (2019): *Fundamentals of Statistics*(7th Revised ed., pp. 2.1-2.21), Hindustan Publishing House.

[10] Duckworth, F. and Lewis, T. "Review of the application of the Duckworth/Lewis method of target resetting in one-day cricket". *Proceedings of the Sixth Australian Conference on Mathematics and Computers* in Sport. Eds: Cohen, G. and Langtry, T. Sydney University of Technology, *NSW*. 1(2002), 27-140, Print.

[11] Lemmer, H.H. "Measures of batting performance in a short series of cricket matches. *South African Statistical Journal*. 42(2008), 83-105, Print

[12] https://numpy.org/case-studies/cricket-analytics/#:~:text=Cricket%20data%20analysis%20requires%20field,of%20data%20cleaning%20and%20preprocessing

[13] https://www.crictracker.com/cricket-analysis/

[14] http://www.iplcricketlive.com/indian-premier-league-news/ipl-brandvalue-at-4-13-billion/

[15] https://timesofindia.indiatimes.com/sports/cricket/ipl/top-stories/ipl-to-give-bcci-95-percent-of-its-surplus/articleshow/62894670.cms

[16] https://internationalcrickethall.com/whats-here/the-game-of-cricket/

[17] www.stats.cricinfo.com

[18] http://www.icc-cricket.com/cricket-rules-and-regulations/

[19] https://www.icc-cricket.com/rankings/mens/player-rankings