SCIENTIFIC JOURNAL OF SPORT AND PERFORMANCE

ORIGINAL ARTICLE

Analysing factors impacting Bangladesh men's T20 cricket performance

Anamul Haque Sajib. *Department of Statisitcs. University of Dhaka. Dhaka, Bangladesh.*
Shakib Hasan Limon. *Department of Statisitcs. University of Dhaka. Dhaka, Bangladesh.*
Adiat Ibn Naser. *Department of Statisitcs. University of Dhaka. Dhaka, Bangladesh.*
Goutam Saha ✉ **.** *Department of Mathematics. University of Dhaka. Dhaka, Bangladesh.*

ABSTRACT

Over the past decade, the T20 cricket format has witnessed a remarkable surge in popularity within the cricketing community. While top-tier teams such as England, Pakistan, Australia, India, and South Africa have embraced this format with ease, others like Bangladesh, Afghanistan, Zimbabwe, and Sri Lanka are facing challenges in adapting to it. This study seeks to identify the key factors influencing the performance of the Bangladesh men's cricket team, utilizing data sourced from ESPNcricinfo. The binary logistic regression model is considered to analyze the data as the response variable is binary (match won or lost). Examining the output of the binary logistic regression, it becomes evident that factors such as 'Location' (the venue where a match took place), 'Most Wickets' (wherein at least one player took 3 or more wickets), and 'One-Digit Score' (signifying scenarios where less than 3, 3 to 6, or more than 6 batsmen scored in the one-digit range) exhibit notable and statistically significant effects on team performance. Notably, variables such as the inclusion of Shakib Al Hasan and the Five Pillars (the concurrent presence of Mashrafe Bin Mortaza, Tamim Iqbal, Shakib Al Hasan, Mushfiqur Rahim, and Mahmudullah Riyad) do not demonstrate any substantial impact on the team's performance. These research findings hold the potential to inform strategic planning aimed at enhancing the Bangladesh men's cricket team's performance.

Keywords: Performance analysis, T20 performance, Bangladesh men's cricket team, ESPNcricinfo data, Binary logistic regression, Performance factors.

Cite this article as:
Sajib, A. H., Limon, S. H., Naser, A. I., & Saha, G. (2024). Analysing factors impacting Bangladesh men's T20 cricket performance. *Scientific Journal of Sport and Performance*, *3*(1), 79-94. https://doi.org/10.55860/ENNL4841

✉ **Corresponding author.** *Department of Mathematics. University of Dhaka. Dhaka-1000, Bangladesh.*
E-mail: gsahamath@du.ac.bd
Submitted for publication November 01, 2023.
Accepted for publication November 21, 2023.
Published December 12, 2023.
Scientific Journal of Sport and Performance. ISSN 2794-0586.

doi: https://doi.org/10.55860/ENNL4841

INTRODUCTION

Cricket, a sport played both indoors and outdoors using a bat and a ball, holds a paramount position as the most cherished and significant sport in Bangladesh (Subhani et al., 2012). The playing area, known as the pitch, is centrally located within a field. Within the framework of specific rules and regulations, a contest unfolds between two teams, each comprising 11 players who take turns in batting and bowling/fielding (Perera, 2015; Singh et al., 2015; Stuelcken et al., 2007). Three different formats are used to play cricket both internationally and domestically, which makes it unique (Petersen *et al*., 2011, Sloane, 2020). These formats include everything from lengthy, multiday Test matches to quick 50 over One-day games and T20 play. The distinct skill sets needed for various formats of cricket are evident through the distinct team compositions chosen for each format in professional cricket (Sloane, 2020). Among these formats, Test cricket stands out as the longest version and is universally acknowledged by coaches, players, and enthusiasts as the ultimate assessment of playing prowess (Wicktramasinghe, 2014, Sloane, 2020). Furthermore, Test cricket is commonly recognized as the zenith of the sport (Peterson et al., 2011; Sloane, 2020). In 1877, during an England tour of Australia, two matches were played against full Australian XIs that are now historically recognized as the inaugural Test matches. The inception of one-day cricket took place in the 1960s, representing a variation of the sport characterized by limited overs, with each side having 50 overs to play (Perera et al., 2015, Sloane, 2020). ODI cricket was introduced as a means to reduce the occurrence of drawn matches and to infuse more excitement into the game through a more aggressive style of batting (Swartz et al., 2006, Sloane, 2020). Subsequently, T20 cricket was introduced with the aim of enhancing the attendance and engagement of contemporary spectators (Singh et al., 2015, Sloane, 2020). The inaugural international T20 match unfolded on June 13, 2005, pitting England against Australia at Hampshire's Rose Bowl. T20 cricket is primarily oriented towards entertaining spectators rather than being a strategic contest of planning and execution (Subhani et al., 2012, Sloane, 2020).

Bangladesh made their T20 International (T20I) debut against Zimbabwe in 2006, initially grappling with the challenges posed by more seasoned teams. Nevertheless, a significant turning point arrived when they clinched their inaugural T20I victory against Scotland during the ICC World T20 in 2007. Over time, Bangladesh began to carve a niche for themselves in the competitive T20 arena, notching up wins against several established cricketing nations. Their spirited displays contributed to a more well-rounded team. In 2014, Bangladesh co-hosted the ICC World T20 with India and managed to progress to the Super 10 stage but couldn't advance further. The 2016 ICC World T20 in India saw Bangladesh deliver a mixed performance. They secured a victory against Oman in the qualifying round but couldn't break through the Super 10 stage. Notably, they have demonstrated their potential by defeating top-ranked teams such as India, Pakistan, and South Africa on multiple occasions.

In Sloane's (2020) report titled "*Analysis of Performance Indicators in IPL Twenty20 Cricket from 2015 to 2017*," data were collected from the reputable source Statsguru, which serves as ESPN Cricinfo's cricket statistics database. The data underwent analysis utilizing the SAS statistical software (SAS, 2017). The dataset was subsequently split into two categories based on whether the team batted first or second. The match outcome was treated as a binary variable, encompassing only win-or-lose results, with drawn matches excluded from the study. To investigate the connection between potential predictor variables and the match outcome, univariate logistic regression was carried out. Each predictor variable was individually incorporated into the model to evaluate its influence on the match result. The statistical significance of each predictor variable was assessed through an exact test, specifically an exact conditional logistic regression, and the resulting precise *p*-values were documented. The entire analysis was executed using the SAS procedure LOGISTIC. In a separate study titled "*IPL Team Performance Analysis: A Multi-Criteria Group Decision*

Approach in a Fuzzy Environment," authored by Dey et al. (2015), data were gathered from the primary dataset available through open sources, covering the six-year period from 2008 to 2013 within the Indian Premier League (IPL). The objective of this report was to assess team performances over the initial six IPL seasons, with a focus on the domain of sports data mining.

In the field of literature, Kimber (1993) introduced a graphical technique for evaluating and contrasting bowlers' performances. Expanding upon this work, Van Staden (2009) introduced another graphical approach specifically designed for comparing the bowling and batting performances of cricketers. This methodology has proven effective in distinguishing various player archetypes, including aggressive batsmen, versatile bowlers, and other notable categories. Dey et al. (2011) put forward a multi-faceted decision-making approach for appraising bowlers' performances within the context of the Indian Premier League. This approach is likely to offer a more comprehensive and nuanced analysis compared to purely graphical methods. Additionally, Barr and Kantor (2004) proposed a mathematical technique for evaluating and selecting batsmen in the sport.

A profound understanding of performance indicators in T20 matches is essential for effective player management and strategic planning in both individual games and tournaments. Coaches and players can gain valuable insights into the impact of various performance variables on match outcomes, enabling them to implement winning strategies. This study aims to provide players and coaches with a deeper comprehension of the factors influencing positive results in matches and tournaments through data collection and analysis.

The study focuses on unravelling the intricate relationship between different aspects of cricket performance and team outcomes in the T20 format for the Bangladesh team. It investigates factors like match location and timing to determine if elements such as home-ground advantage or specific time slots influence their success. Additionally, the study delves into batting performance variables to highlight the importance of not only the team's overall run tally but also individual batsmen's roles and the strategic decision of batting first. In the domain of bowling performance, the study aims to assess how effectively the bowlers' wicket-taking abilities contribute to the team's victories or defeats. Ultimately, this comprehensive analysis aims to offer a holistic understanding of the factors shaping the Bangladesh cricket team's performance in T20 matches, providing valuable insights for cricket enthusiasts and strategists.

METHODOLOGY

Data and variables

The Bangladesh National Team has participated in a total of 142 T20 matches since 2006. Due to the relatively limited number of T20 matches played by the Bangladesh national team, all 142 matches have been considered as in our dataset. Within this dataset, we have a single dependent variable referred to as 'response.' The 'response' variable signifies the match outcome, distinguishing between a win (1) and a loss (0).

Our study incorporates eleven explanatory variables, all of which are categorical.

The first explanatory variable, 'Location,' identifies the match's location, categorized into three options: '0' for matches played at home, '1' for matches played in Asian countries other than Bangladesh, and '2' for matches played in countries outside Asia.

The second explanatory variable, 'Time,' contains two categories, representing whether the match occurred during the daytime (0) or nighttime (1).

We also consider the presence or absence of Shakib Al Hasan, a prominent all-rounder, as an independent variable ('Shakib'). If Shakib played, it is denoted as '1,' and if not, it is '0.'

The 'Five Pillars' explanatory categorical variable reflects the participation of the five key players, Mashrafe Bin Mortaza, Tamim Iqbal, Shakib Al Hasan, Mushfiqur Rahim, and Mahmudullah Riyad together. It takes on '1' when all five members were present and '0' if at least one member was absent during the match.

The 'Captain' variable indicates the captaincy role during the match, with '0' for Mahamudullah, '1' for Mashrafeee, '2' for Shakib, and other Bangladeshi players.

'Batting First' is the sixth explanatory variable, indicating whether Bangladesh batted in the first innings (1) or the second innings (2).

The 'Score' independent variable categorizes matches into '0' if Bangladesh scored 150 runs or less and '1' if they scored more than 150 runs.

'One Digit Score' has three categories: '0' for matches where fewer than 3 batsmen scored one-digit runs, '1' for matches with 3 to 6 batsmen scoring one-digit runs, and '2' for matches with more than 6 batsmen scoring one-digit runs.

The 'Most wicket' qualitative independent variable indicates whether at least one bowler took 3 or more wickets (1) or not (0).

The 'Thirty-plus run' variable signifies whether at least two players scored thirty-plus runs (1) or not (0).

Lastly, the 'Boundaries' variable distinguishes matches where Bangladesh scored 20 or fewer boundaries (0) from those where they scored more than 20 boundaries.

Study variables

Table 1 displays the dependent and independent variables used in this study, along with their respective categories.

Table 1. Dependent and independent variables along with categories.

Level of Variables
Dependent Variable
Response (0: Lose, 1: Win)
Independent Variables
Location (0: Home, 1: Asia, 2: Other)
Time (0: Day, 1: Night)
Shakib (0: Absent, 1: Present)
Five Pillars (0: Absent, 1: Present)
Captain (0: Mahmudullah, 1: Mashrafee, 2: Shakib, 3: Others)
Bat First (0: No, 1: Yes)
Score (0: Match score equal or less than 150, 1: Match score more than 150)

One-digit score (0: Less than 3 batsmen scored 1 digit, 1: One digit scored by 3 to 6 batsmen, 2: More than 6 batsmen scored one digit)
Most wickets (1: At least 1 player has taken 3 or more wickets, 0: Otherwise)
Thirty plus run (1: At least 2 players scored 30+ runs, 0: Otherwise)
Boundaries (0: Boundaries scored less than 20, 1: Boundaries scored more than 20)

Method of data collection

Bangladesh has been actively participating in T20 matches since 2006. Our research encompasses data collected from every T20 match played by the Bangladesh National Team spanning the years 2006 to 2023. All the data utilized in this study has been sourced from the ESPNcricinfo website, which is renowned for maintaining comprehensive match data, including live ball-by-ball commentary. ESPNcricinfo is widely recognized as a dependable source, trusted not only within professional cricket but also referenced by numerous published authors.

Data processing

Following the input of data from ESPNcricinfo, the dataset underwent comprehensive analysis through the utilization of IBM SPSS Statistics software. All pertinent characteristics within the dataset were scrutinized, organized, summarized, and assessed in accordance with the goals of the research.

Statistical analysis

In our analysis, we conducted a descriptive statistical analysis of the study variables, utilizing frequencies and percentages to elucidate their characteristics. To explore the connection between categorical outcomes and categorical explanatory variables, we applied either the Pearson Chi-Square test or Fisher's exact test. Furthermore, Binary Logistic Regression was employed as a part of our analytical approach.

Univariate analysis

Since all variables considered in this study are categorical variables, we calculated the frequency and percentage distribution of each covariate.

Bivariate analysis

To examine how the response of the match (win or lose) changes with the change of category of each covariate we considered, we used the Pearson Chi-square test.

The hypothesis in bivariate analysis can be formulated as:

H_0: There is no association between the selected covariate and the outcome variable. H_a: There is an association between the selected covariate and the outcome variable.

The Chi-square test statistic can be defined as:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

where O_{ij}= The number of observations in the cell (i, j); $i = 1,2,3 \ldots, r, j = 1,2,3, \ldots, c$; E_{ij}: Expected cell values. The test statistic follows a Chi-squared distribution with $(r-1) * (c-1)$ degrees of freedom. The *p*-value collected from this test is used to make the decision.

Logistic regression

Logistic regression is a statistical analysis method of modelling the probability of a discrete outcome given input variables. The formula of logistic regression can be written in a linear equation form:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k,$$

where p=Pr(Y=1) and $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are the regression coefficients and $X_1, X_2, \ldots, X_k$ are explanatory variables. Solving for probability equation result in:

$$p = \Pr(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)}}$$

Binary logistic regression

The most popular type of logistic regression model is an outcome that can only have one of two possible values, such as yes or no, win or lose, true or false, and so on. The response variable can only fall into one of two categories in binary logistic regression.

Issues during data collection

Manual data collection from websites can prove to be a laborious endeavour, particularly when grappling with substantial datasets or the need to extract information from numerous web pages. The complexity intensifies as websites frequently present data in assorted formats, creating hurdles in the seamless standardization and integration of data into statistical analysis tools like SPSS. Moreover, the reliability of web-sourced data can be compromised as certain websites might lack comprehensive information or omit vital variables, such as runs and the number of wickets fallen during the power play, thereby limiting the depth of analysis. Additionally, manual data entry is susceptible to human errors, encompassing typographical mistakes and misinterpretation of data, which, in turn, can undermine the precision and trustworthiness of subsequent analyses.

RESULTS AND DISCUSSION

Univariate analysis

In this section, we have evaluated the data by examining each variable separately, which is the simplest form to analyse the data. To do this, we have used a univariate Table 2 to display the frequency distributions and corresponding percentages of our explanatory variables as well as the response variable.

Table 2. Percentage and frequency distribution of variables with different categories.

Variable	Frequency	Percentage	Valid Percentage
Response			
Win	55	38.7	38.7
Lose	87	61.3	61.3
Location			
Home	60	42.3	42.3
Asia	35	24.6	24.6
Other	47	33.1	33.1
Time			
Day	64	45.1	45.1

Night	78	54.9	54.9
Shakib			
Present	33	23.2	23.2
Absent	109	76.8	76.8
Five pillars			
Present	116	81.7	81.7
Absent	26	18.3	18.3
Captain			
Mahmudullah	39	27.5	27.5
Mashrafe	24	16.9	16.9
Shakib	37	26.1	26.1
Other	42	29.6	29.6
Bat First			
No	71	50	50
Yes	71	50	50
Score			
<150	90	63.4	63.4
>150	52	36.6	36.6
One Digit score			
<3 players	48	33.8	33.8
3 to 6 players	54	38	38
>6 players	40	28.2	28.2
Most wicket			
At least 1 player has taken 3 or more wickets	9	6.3	6.3
otherwise	133	93.7	93.7
Thirty plus run			
At least 2 players scored 30+ runs	15	10.6	10.6
otherwise	127	89.4	89.4
Boundaries			
Boundaries scored less than 20	114	80.3	80.3
Boundaries scored more than 20	28	19.7	19.7

The Bangladesh cricket team has participated in a total of 142 T20 matches, winning 55 of them, which translates to a winning percentage of 38.7%. Conversely, their losing percentage stands at 61.3%. Bangladesh team has played most of their matches in their home country which is 42.3% of the total matches while 24.6% of the matches have been played within Asia (outside Bangladesh) which is the lowest.

Out of the 142 matches, 78 took place during nighttime, while 45.1% were daytime fixtures. This indicates that Bangladesh played 9.8% more T20 matches in nighttime conditions.

The "*5 pillars*" of the Bangladesh cricket team, comprising Tamim Iqbal, Shakib Al Hasan, Mushfiqur Rahim, Mashrafe, and Mahmudullah, collectively featured in only 26 out of the 142 matches. In a majority of the T20 matches, at least one member of this quintet was absent, accounting for 81.7% of the games. Conversely, all five players were present in a mere 18.3% of the matches.

Mashrafe held the position of team captain in the lowest proportion of matches, accounting for just 16.9%. In contrast, Shakib and Mahmudullah led the team in 26.1% and 27.5% of the matches, respectively. Notably, Bangladeshi players other than Mashrafe, Shakib, and Mahmudullah assumed the captain's role in the highest proportion of matches, totalling 29.6% within the group.

In half of the T20 matches, the Bangladesh squad took to the crease for the 1st innings, while the remaining 50% saw them batting in the 2nd innings. In 90 out of the 142 matches, the Bangladesh team's scores fell below the 150-run mark, signifying a rate of 36.6%. Conversely, in 36.6% of the matches, the Tigers managed to amass more than 150 runs.

In the lowest percentage of T20 matches, which amounts to 28.2%, more than six players scored in single digits. In 54 matches, which is 38% of the total, three to six players registered single-digit scores. Finally, in 33.8% of the matches, fewer than three Bangladeshi players scored in single digits.

In a mere 9 T20 matches, there was the presence of at least 1 Bangladeshi bowler who managed to secure 3 or more wickets. This occurrence constituted just 6.30% of all matches. Conversely, a significant majority of matches, precisely 93.7%, did not witness any bowler achieving 3 or more wickets.

In 15 matches, at least 2 players scored 30+ runs, reflecting a rate of occurrence of 10.6%. Conversely, in 89.4% of the matches, fewer than 2 players struggled to reach the 30-run mark.

The Bangladesh team successfully notched up more than 20 boundaries (comprising 4's and 6's) in 19.7% of the matches. In contrast, in the majority of T20s, specifically 80.3% of them, which amounts to 114 matches, the Bangladesh squad fell short of this boundary count.

Bivariate analysis

In this section, we conducted a bivariate analysis to examine the relationships between our independent variables and the match result. To achieve this, we initially created a bivariate Table. Significant associations can be discerned from the *p*-values presented in Table 3. Table 3 displays the frequencies and percentages of various explanatory variables along with their corresponding *p*-values.

The analysis revealed that the 'Location,' 'Score,' 'One-digit score,' 'Most wicket,' '30+ run,' and 'Boundaries' variables significantly influence the match outcome, as their respective *p*-values are less than the threshold $\alpha = .05$.

Furthermore, Table 3 indicates that no significant associations exist between the timing of the match, the presence of Shakib, the presence of the five key players (i.e., Tamim, Shakib, Mushfique, Mahamudullah, Mashrafe), the captaincy, and batting during the first innings.

Table 3 shows that response (i.e. win or lose) differs by the place where the game was played. The home ground plays a significant role in the chance of winning or losing.

In instances where the match was played away from home, the winning percentage experiences a decline. This observation is underscored by the low *p*-value of .001, further highlighting the significance of this trend. When Bangladesh's score falls below 150, they secure victory in only 31.1 percent of the matches. Conversely, this figure rises to 51.9 percent when they score above 150. Our statistical significance test

corroborates this finding, as the *p*-value of .012 is less than the threshold α = .05. Therefore, this conclusion holds significance at the 5% level of significance.

Table 3. Percentage and frequency distribution of variables with different categories of selected covariates along with *p*-values.

Explanatory Variables	Win	Lose	*p*-value
Location			
Home	33 (55%)	27 (45%)	
Asia	10 (28.6%)	25 (71.4%)	.003**
Other	12 (25.5%)	35 (74.5%)	
Time			
Day	21 (32.8%)	43 (67.2%)	.227
Night	34 (43.6%)	44 (56.4%)	
Shakib			
Present	44 (40.4%)	14 (53.8%)	.505
Absent	11 (33.3%)	73 (62.9%)	
5 Pillars			
Present	12 (46.2%)	14 (53.8%)	.505
Absent	43 (37.1%)	73 (62.9%)	
Captain			
Mahmudullah	14 (35.9%)	25 (64.1%)	
Mashrafee	10 (41.7%)	14 (58.3%)	.963
Shakib	14 (37.8%)	23 (62.2%)	
Other	17 (40.5%)	25 (59.5%)	
Bat First			
No	24 (33.8%)	47 (66.2%)	.301
Yes	31 (43.7%)	40 (56.3%)	
Score			
<150	38 (31.1%)	62 (68.9%)	.012*
>150	27 (51.9%)	25 (48.1%)	
One Digit score			
<3 players	30 (62.5%)	18 (37.5%)	
3 to 6 players	19 (35.2%)	35 (64.8%)	0***
>6 players	6 (15.0%)	34 (85.0%)	
Most wicket			
At least 1 player has taken 3 or more wickets	8 (88.9%)	1 (11.1%)	.002**
otherwise	47 (35.3%)	86 (64.7%)	
Thirty plus run			
At least 2 players scored 30+ runs	10 (66.7%)	5 (33.3%)	.025*
otherwise	45 (35.4%)	82 (64.6%)	
Boundaries			
Boundaries scored less than 20	39(34.2%)	75(65.8%)	.031*
Boundaries scored more than 20	16(57.1%)	12(42.9%)	

*Note. *p < .05, **p < .01, ***p < .001.*

Furthermore, the match outcome is influenced by the number of players who score in single digit. As the count of players scoring in single digit increases, the winning percentage declines. For example, when fewer than 3 players score in single digits, the team wins 62.5% of the matches. When the count exceeds 3 but remains below 6, the winning percentage drops to 35.2%, and when it surpasses 6, it diminishes further to 15%. Additionally, the *p*-value is 0, indicating a highly significant relationship at the 5% level of significance.

It is evident that the number of wickets taken by a single bowler also affects the match outcome. The higher the wicket count, the greater the likelihood of winning. When at least one player takes 3 or more wickets, the team wins in 88.9% of cases. The *p*-value is .002, which is less than $\alpha = .05$, demonstrating a significant association at the 5% level of significance as well.

The number of batsmen scoring more than 30 runs evidently impacts the team's chances of winning. As this number increases, the winning percentage also rises. When at least 2 players score more than 30 runs, the Bangladesh team tends to win 66.7% of the matches. This association is statistically significant at the 5% level, as indicated by the *p*-value of .025, which is less than $\alpha = .05$.

Similarly, the number of boundaries scored has a noticeable effect on the team's likelihood of winning. As the boundary count increases, the winning percentage (57.1%) surpasses that when the count is less than 20. This relationship is statistically significant at the 5% level, with a *p*-value of .031, which is less than $\alpha = .05$.

Binary logistic model

In this section, we have done our required analysis on the basis of Binary Logistic Regression. In Binary Logistic Regression analysis, we have selected the first category of response as the baseline category which is 'Lose'. Similarly, the first category of each of our categorical explanatory variables has been considered as the baseline category.

Location

We employed 'Home' as the reference category for the 'Location of the match' variable in our analysis. We then compared the odds ratios for the remaining two categories: 'Asia' and 'Other region.' According to the results presented in Table 4, the odds ratio for the 'Asia' category is 0.244. This implies that the odds of winning a match are approximately 0.244 times lower when the match is played in Asia (away from home) compared to matches played at home. In percentage terms, this suggests a 75.59% reduction in the likelihood of winning when the match is held in Asia as opposed to home.

Likewise, for the 'Other region' category, the odds ratio is 0.245, indicating that the odds of winning a match are approximately 0.245 times when the match is played in other regions compared to matches played at home. In percentage terms, this translates to a 75.47% decrease in the chance of winning when the match occurs in other regions rather than at home. The corresponding *p*-values for both categories are .015845 and .013582, respectively. These *p*-values fall below the conventional significance level of .05, signifying that the coefficients for both 'Asia' and 'Other region' are statistically significant at the 5% significance level. As a result, we have sufficient evidence to reject the null hypothesis and conclude that the location of the match significantly impacts the likelihood of winning the match.

Time

We've designated 'Day' as the baseline category for the 'Time' covariate in our analysis. As shown in Table 4, the odds ratio for 'The match is played at night' is 1.082. This odds ratio indicates that the likelihood of

winning the match at night is approximately 1.082 times that of winning during the day. Alternatively, we can express this as an 8.2% higher probability of winning the match when it's played at night compared to daytime. However, since the associated *p*-value exceeds the conventional significance threshold of 0.05, we lack sufficient evidence to deem the 'Time' coefficient statistically significant at the 5% significance level. Consequently, we cannot confidently assert that the timing of the match (day or night) significantly influences the odds of winning, based on our data analysis.

Shakib

We chose 'Absent' as the baseline category for the 'Shakib' variable in our analysis. According to the findings in Table 4, the odds ratio for 'Shakib being present in the match' stands at 1.1668. This odds ratio signifies that the likelihood of winning the match is approximately 1.1668 times higher when Shakib is present compared to when he is absent. In simpler terms, the presence of Shakib in the match is associated with a 16.68% greater chance of winning than when he is absent. However, the *p*-value linked with this odds ratio is .801, surpassing the customary significance threshold of .05. This result indicates that the 'Shakib' coefficient does not attain statistical significance at the 5% level. Consequently, we lack sufficient evidence to reject the null hypothesis, which suggests that Shakib's presence does not significantly impact the likelihood of winning the match.

Five pillars

We have designated 'absent' as the baseline category for the 'Five Pillars' covariate in our analysis. As per the findings presented in Table 4, the odds ratio for the 'Five Pillars' being 'present' stands at 0.867. This odds ratio implies that the odds of winning the match are approximately 0.867 times when the 'Five Pillars' are present compared to when they are absent. To put it differently, if the 'Five Pillars' are present, there is a 13.3% reduced chance of winning the match compared to when they are absent. However, the associated *p*-value for this odds ratio is .829, exceeding the customary significance threshold of .05. Consequently, the coefficient for 'Five Pillars' does not achieve statistical significance at the 5% level. Therefore, we lack sufficient evidence to reject the null hypothesis, which suggests that the presence of the 'Five Pillars' does not exert a statistically significant influence on the odds of winning the match, based on the data analysed.

Captain

We selected 'Mahmudullah' as the reference category for the 'Captain' variable in our analysis. We then compared the odds ratios for the three other categories: 'Mashrafe,' 'Shakib,' and 'Others.' According to the results presented in Table 4, the odds ratio for the 'Mashrafe' category is 0.976. This implies that the odds of winning a match are approximately 0.976 times when Mashrafe is the captain, in comparison to Mahmudullah being the captain. This translates to a 2.4% lower chance of winning the match when Mashrafe is the captain. Similarly, for the 'Shakib' category, the odds ratio is 1.368. This suggests that the odds of winning a match are approximately 1.368 times higher when Shakib is the captain compared to Mahmudullah being the captain, corresponding to a 36.8% greater chance of winning. Additionally, for the 'Others' category, the odds ratio is 1.93, indicating that the odds of winning a match are approximately 1.93 times when the captain is from another category compared to Mahmudullah being the captain. This results in a 93% higher chance of winning the match.

However, the associated *p*-values for 'Mashrafe,' 'Shakib,' and 'Others' are 0.97, 0.64, and 0.35, respectively. These *p*-values exceed the conventional significance level of .05, indicating that the coefficients for 'Mashrafe,' 'Shakib,' and 'Others' are not statistically significant at the 5% level. Therefore, we lack sufficient evidence to reject the null hypothesis, suggesting that the choice of captain from the 'Mashrafe,' 'Shakib,' or

'Others' categories does not exert a statistically significant impact on the likelihood of winning the match compared to Mahmudullah being the captain.

Bat first

We have designated 'NO' as the reference category for the 'Bat First' covariate in our analysis. According to the findings presented in Table 4, the odds ratio for 'Bat First' being 'YES' is 1.588. This odds ratio signifies that the odds of winning the match are approximately 1.588 times when the Bangladesh team opts to bat first compared to when they do not choose to bat first. Alternatively, we can express this as a 58.88% increased chance of winning the match when the Bangladesh team decides to bat first. However, the associated *p*-value for this odds ratio is .293, which surpasses the customary significance level of .05. Consequently, the coefficient for 'Bat First' does not achieve statistical significance at the 5% level. Therefore, we lack sufficient evidence to reject the null hypothesis, implying that the decision to bat first by the team does not have a statistically significant impact on the likelihood of winning the match, based on the data analysed.

Table 4. Coefficients, Standard Errors, Odds Ratios and p values of the Binary Logistic Regression.

Explanatory Variables	Coefficient (β)	Standard Error	Odds ratio (e^{β})	*p*-value
Intercept	0.506	0.778	1.659	.515
Location				
Asia	-1.41	0.585	0.244	.016
Other	-1.405	0.569	0.245	.014
Time				
Night	0.079	0.497	1.082	.874
Shakib				
Present	0.154	0.614	1.167	.802
Five Pillars				
Present	-0.143	0.663	0.867	.829
Captain				
Mashrafe	-0.024	0.735	0.976	.974
Shakib	0.313	0.676	1.368	.643
Other	0.657	0.702	1.93	.349
Bat First				
Yes	0.462	0.44	1.588	.293
Score				
>150	0.098	0.583	1.103	.867
One Digit score				
3 to 6 players	-1.133	0.495	0.322	.022
>6 players	-2.726	0.709	0.065	0
Most wicket				
At least 1 player has taken 3 or more wickets	3.168	1.304	23.749	.015
Thirty plus run				
At least 2 players scored 30+ runs	0.005	0.753	1.005	.995
Boundaries				
Boundaries scored more than 20	0.111	12(42.9%)	1.117	.869

Score: We have designated 'Less than 150' as the reference category for the 'Score' covariate in our analysis. We then compared the odds ratio for the 'More than 150' category. According to the findings presented in Table 4, the odds ratio for 'Score' being 'More than 150' is 1.103. This odds ratio suggests that the odds of winning the match are approximately 1.103 times when the team's score is 'More than 150' compared to when it is 'Less than 150.' In percentage terms, this translates to a 10.3% greater chance of winning the match when the team's score exceeds 150 compared to when it is below 150. However, the corresponding *p*-value for 'Score' being 'More than 150' is .867, exceeding the conventional significance level of .05. This result indicates that the coefficient for 'Score' is not statistically significant at the 5% level. Consequently, we lack sufficient evidence to reject the null hypothesis, suggesting that the difference in match outcomes based on the team's score is not statistically significant at the 5% level.

One-Digit Score: In our analysis, we explored the connection between the 'One-digit score' covariate and the likelihood of winning a cricket match. The 'One-digit score' variable encompasses three categories: 'Less than three players' (which we use as the reference category), 'Between three to six players,' and 'More than six players.' According to the findings presented in Table 4, for the 'Between three to six players' category, the odds ratio is 0.322. This indicates that when there are between three to six players who score one-digit runs, the chances of winning the cricket match are approximately 0.322 times compared to when there are fewer than three players scoring one-digit runs. In simpler terms, having between three to six players with one-digit scores is associated with a 67.8% reduced likelihood of winning the match compared to having less than three players with one-digit scores. Furthermore, the corresponding *p*-value for this odds ratio is .022, falling below the significance threshold of .05. This suggests that the coefficient for 'Between three to six players' is statistically significant, and we can confidently affirm that this category has a significant impact on the outcome of winning the cricket match.

For the 'More than six players' category, the odds ratio is 0.065. This implies that when there are more than six players with one-digit scores, the odds of winning the match are approximately 0.065 times compared to when there are fewer than three players with one-digit scores. In simpler terms, having more than six players with one-digit scores is associated with a substantial 93.5% reduction in the odds of winning the match compared to having less than three players with one-digit scores. Additionally, the corresponding *p*-value for this odds ratio is .0001, significantly below the significance level of .05. This indicates that the coefficient for 'More than six players' is highly statistically significant, and we can confidently assert that this category exerts a significant influence on the outcome of winning the cricket match.

Most wicket

We have defined the category of interest for the 'Most wicket' covariate as 'At least 1 player has taken 3 or more wickets,' while the reference category pertains to situations where no player has achieved this milestone in our analysis. According to the findings presented in Table 4, the odds ratio for 'At least 1 player has taken 3 or more wickets' is 23.749. This odds ratio signifies that when at least one player accomplishes this feat, the odds of winning the match increase by approximately 23.75 times compared to situations where no player has achieved it. In simpler terms, if the team manages to have at least one player take 3 or more wickets, there is a substantial and statistically significant improvement in the likelihood of winning the match.

Additionally, the *p*-value associated with this odds ratio is .015152, which falls below the conventional significance level of .05. This implies that the coefficient for 'At least 1 player has taken 3 or more wickets' is statistically significant at the 5% level. Consequently, we have robust evidence to reject the null hypothesis, firmly asserting that having at least one player who has taken 3 or more wickets significantly influences the team's chances of winning the match.

Thirty plus run

We have selected 'At least 2 players scored 30 plus runs' as the baseline category for the 'Thirty plus run' covariate in our analysis. We then compared the odds ratio for the alternative category, denoted as 'otherwise.' Based on the findings presented in Table 4, the odds ratio for 'otherwise' is 1.005. This odds ratio signifies that the odds of winning the match are approximately 1.005 times when there are fewer than 2 players who scored 30 plus runs (i.e., 'otherwise') in comparison to the scenario where at least 2 players scored 30 plus runs. In terms of percentages, this translates to a slight 0.5% increase in the likelihood of not winning the match when fewer than 2 players achieve this run milestone, as opposed to when at least 2 players score 30 plus runs. However, it's worth noting that the *p*-value exceeds the conventional significance level of .05, indicating that the coefficient for 'Thirty plus run' is not statistically significant at the 0.05% level.

Boundary

We have designated 'Boundaries scored less than 20' as the baseline category for our analysis. According to the data presented in Table 4, the odds ratio for 'Boundaries scored more than 20' in the match is 1.117. This implies that the odds of winning the match increase by a factor of 1.117 when there are 'Boundaries scored more than 20' in comparison to when 'Boundaries scored less than 20' are observed. Put differently, if the team manages to score 'Boundaries more than 20,' there is an 11.7% higher chance of winning the match compared to when they score 'Boundaries less than 20.'

However, the associated *p*-value of .869 indicates that the coefficient is not statistically significant at the 0.05% significance level. Consequently, we lack sufficient evidence to reject the null hypothesis, suggesting that 'Boundaries scored less than 20' does not have a significant impact on the likelihood of winning the match.

CONCLUSIONS

In our comprehensive research on Bangladesh national T20 cricket from 2006 to 2023, we uncovered key insights into the factors behind successful performances.

1. In the last decade, T20 cricket has surged in popularity, captivating audiences and players worldwide. While some cricket powerhouses like England, Pakistan, Australia, India, and South Africa have embraced the format, others, including Bangladesh, Afghanistan, Zimbabwe, and Sri Lanka, have faced challenges adapting to it.
2. Our research aimed to analyse factors affecting Bangladesh men's cricket team performance using ESPN Cricinfo data. We used binary logistic regression to study whether matches were won or lost.
3. Our findings were intriguing. Three key factors stood out among the variables examined: "*Location*" of the match, the performance of a player with "*Most Wickets*," and the frequency of "*One-Digit Scores*" all had significant impacts on the team's overall performance.
4. Surprisingly, expected factors like Shakib and the "*Five Pillars*" had no significant impact on the team's performance.
5. Our findings have profound implications, providing a strategic roadmap for the Bangladesh men's cricket team to improve their T20 performance and shape their cricketing future.

Practical applications

1. Match location significantly impacts Bangladesh Team's performance. Adapting strategies, pitch preparation, and team compositions for different locations can enhance their chances of winning away from home.

2. The strong correlation between wickets taken by bowlers and positive match outcomes underscores the importance of a potent bowling attack. Focusing on bowlers' skills, strategic variations, and match-specific combinations can consistently disrupt opponent strategies.
3. The correlation between runs scored by batsmen and positive match outcomes highlights the need for focused training in innovative stroke play and countering various bowling styles, ensuring consistent and impactful contributions from the batting unit.

Limitations

Our analysis mainly centred on specific factors like match location, match time, presence of Shakib, captaincy, batting, and bowling. We didn't delve into other potentially influential variables such as weather, player form, injuries, fielding, team morale, captaincy decisions, and opposition strength, which could provide a more comprehensive understanding of performance dynamics. Limited available research and data constraints led to the exclusion of various performance variables from our study.

AUTHOR CONTRIBUTIONS

Conceptualization: Anamul Haque Sajib. Supervision: Anamul Haque Sajib. Methodology: Anamul Haque Sajib, Shakib Hasan Limon, Adiat Ibn Naser. Data curation: Shakib Hasan Limon, Adiat Ibn Naser. Writing—original draft preparation: Anamul Haque Sajib, Shakib Hasan Limon, Adiat Ibn Naser, Goutam Saha. Writing—review and editing: Anamul Haque Sajib, Shakib Hasan Limon, Adiat Ibn Naser, Goutam Saha. All authors have read and agreed to the published version of the manuscript.

SUPPORTING AGENCIES

No funding agencies were reported by the authors.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

REFERENCES

Bangladesh Cricket Board. (2021). Official T20 International Records. Retrieved from [Accessed 12/05/2023]: https://www.tigercricket.com.bd/

CricketStats+. (2022). Advanced Cricket Statistics and Analytics.

Dey, P. K., Ghosh, D. N., & Mondal, A. C. (2011). A MCDM approach for evaluating bowlers performance in IPL. Journal of emerging trends in Computing and Information Sciences, 2(11).

Dey, P. K., Ghosh, D. N., & Mondal, A. C. (2015). IPL team performance analysis: a multi-criteria group decision approach in fuzzy environment. IJ Information Technology and Computer Science, 8, 8-15. https://doi.org/10.5815/ijitcs.2015.08.02

ESPNCricinfo. (2021). Bangladesh Cricket Team's Profile. Retrieved from [Accessed 12/05/2023]: https://www.espncricinfo.com/team/bangladesh

Kimber, A. (1993). A graphical display for comparing bowlers in cricket. Teaching Statistics, 15(3), 84-86. https://doi.org/10.1111/j.1467-9639.1993.tb00664.x

Kimber, A. C., & Hansford, A. R. (1993). A statistical analysis of batting in cricket. Journal of the Royal Statistical Society Series A: Statistics in Society, 156(3), 443-455. https://doi.org/10.2307/2983068

Perera, G.H. P (2015). Cricket Analytics.

Petersen, C., Pyne, D. B., Portus, M. J., & Dawson, B. (2008a). Analysis of Twenty/20 cricket performance during the 2008 Indian Premier League. International Journal of Performance Analysis in Sport, 8(3), 63-69. https://doi.org/10.1080/24748668.2008.11868448

Petersen, C., Pyne, D. B., Portus, M. R., Cordy, J., & Dawson, B. (2008b). Analysis of performance at the 2007 Cricket World Cup. International Journal of Performance Analysis in Sport, 8(1), 1-8. https://doi.org/10.1080/24748668.2008.11868417

Sharma, S. K. (2013). A factor analysis approach in performance analysis of T-20 cricket. Journal of Reliability and Statistical Studies, 69-76.

Singh, S. (2011). Measuring the performance of teams in the Indian Premier League. American Journal of Operations Research, 1(3), 180. https://doi.org/10.4236/ajor.2011.13020

Sloane, S. (2020). Analysis of performance indicators in IPL Twenty20 Cricket from 2015 to 2017 (Doctoral dissertation, University of the Free State).

Subhani, M. I., Hasan, S. A., & Osman, M. A. (2012). Will T20 clean sweep other formats of Cricket in future?.

Van Staden, P. J. (2009). Comparison of cricketers' bowling and batting performances using graphical displays.

Wickramasinghe, I. P. (2014). Predicting the performance of batsmen in test cricket. https://doi.org/10.14198/jhse.2014.94.01

