Kurdish Studies
August 2024
Volume: 12, No: 5, pp 781-788
ISSN: 2051-4883 (Print) | ISSN 2051-4891 (Online)
www.KurdishStudies.net

DOI: 10.53555/ks.v12i5.3334

Analyzing the Impact of Venue and Match Factors on Ground-Wise Average Scores and Successful Run Chases in T-20I Cricket Using Factorial Design

Syed Asghar Ali Shah[1*], Qamruz Zaman[2], Sundas Hussain[3], Soofia Iftikhar[3], Syed Habib Shah[4], Abdur Rehaman Sabir[1], Sofia Gul[5], Neelam[1], Mansoor Ahmad[1], Abbas Ali, [1]Aizaz Shah[1]

[1]Cereal Crops Research Institute (CCRI), Pirsabak, Nowshera
[2] Department of Statistics, University of Peshawar, Pakistan
[3]Shaheed Benazir Bhutto Women University Peshawar, Pakistan
[4] Institute of Numerical Sciences, Kohat University of Science and Technology, Pakistan
[5]College of Home Economics, University of Peshawar, Pakistan

Abstract
The purpose of this study is to evaluate the impact of four key factors i.e. venue, match conditions, toss outcomes, toss decisions on ground-specific highest run chases in Twenty20 international cricket (T-20I) using a full factorial design. The study uses a stepwise regression model to analyze match data from Pakistan, Australia, India, and New Zealand in order to find significant influences on average scores pursued. The findings show that toss decision (TD) and its interactions with venue (Vn) and match condition (MC) are important predictors of run chase outcomes; the final model accounts for 66.8% of the variation in scores. To improve the predicted accuracy of the model, future study ought to consider about incorporating more variables like team makeup, individual player performance, and weather etc.

Key words: Venue, T-20I, Match conditions, Factors, Interactions, factorial design, outcomes, Toss decision

1. Introduction

T-20 International cricket has brought a new dimension to the world of cricket where each ball and run makes a difference. Most notably, bowlers and batsmen have to be aggressive due to the nature of T-20I played under a very flexible format of 20 overs per team where players need to be more creative and intense. This format has some of the most attractive aspects of cricket like teams having to chase down totals set by the opposition through critical quick scoring. Foretelling the highest successful run chases in T-20 cricket is no doubt a very tough and important task because it plays a very vital role in driving the strategy and decision-making of a particular team.

Many factors have been seen to affect run chases in T-20I cricket these include; pitch condition, ground where the game is being played, outcome of the toss, position of players and even the prowess of the two teams. These factors do not work in isolation; instead, there exists an intricate relationship which makes it impossible to forecast the result of the run chase. To counter this, factorial design, a statistical tool that facilitates the study of multiple variables and their interactions offers a good starting line in modeling and analyzing the results of run chases. Many researchers have applied the factorial design in different discipline for the purpose of optimizing responses, making prediction where many factors affect the result. In the context of T-20I cricket to understand the impact of various factors such as match environment, match venue, toss outcome and toss decision for run chases the factorial design can be applied. When these factors are adjusted in a structured manner, it becomes easy for the researchers to understand which of the variables is most influential and how they combine to influence the outcome [1].

The link between factors and the outcomes using design of experiment (DOE) approach [2] as depicted in Figure-1;

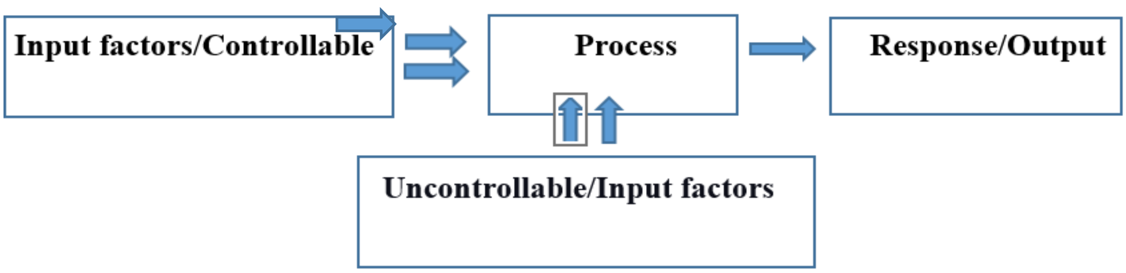

Figure 1: DOE Procedure for Input and Output Factors

Factorial design is a relatively new approach in sports analytics, with various benefits, especially in cricket. In order to provide a more thorough understanding of the variables at play, it first enables the efficient study of several factors at once. Second, it helps to identify important interactions between variables, which might be important information to have when attempting to make precise predictions. Finally, researchers can create testable and validated predictive models through the use of factorial design, which increases the predictability of the results. [3]

In order predict the high average ground specific run chases in Twenty20 International Cricket, this study will make use of the factorial design. This study aims to identify the crucial elements and relationships that lead to effective run chases by examining a dataset that contains information on match conditions, venue features, toss outcomes, and toss decision.

2. Literature Review

Previous studies have emphasized the significance of a number of variables in influencing cricket match outcomes. As per Clarke and Norman [4] and Smith and Jones [5], research has indicated that venue attributes and pitch conditions have a noteworthy impact on the result of matches. Kumar and Agarwal [6] have highlighted the toss outcome and team rankings as crucial factors that determine the success of a match. The relationship between these variables and how they work together to affect run chase success in Twenty20 international cricket, however, is still little known. This work aims to close this gap by analyzing these interactions and creating predicting models for successful run chases using factorial design.Due to the format's complexity and quick pace, research on the variables affecting run chases in Twenty20 International (T-20I) cricket has attracted a lot of attention lately. In this field, research has aimed to pinpoint the critical elements that influence the results of matches, with an emphasis on the circumstances surrounding successful run chases. In-depth study of one-day cricket was conducted by Clarke and Norman[4], who found that weather, team strength, and pitch conditions were important variables in determining the outcome of matches. Even though the focus of their study was on the one-day format, given the similarity in the elements at play, the observations apply to T-20I cricket as well. However, the compressed format of T-20I matches increases the influence of these factors, therefore it is crucial to comprehend how they interact in order to make reliable predictions about the results.

The impact of venue features on T-20 cricket match outcomes has been the subject of additional research. According to Smith and Jones [5], ground dimensions, altitude, and local weather all play a significant role in creating unique playing settings that can significantly impact the outcome of a run chase. For example, shorter boundary lines encourage high-scoring games more than higher ones, which might have an impact on player strength and ball movement. These venue-specific elements are critical to consider while studying run chases because, depending on the circumstances of the match, they can either give the chasing team an advantage or create a difficulty. To create reliable prediction models for Twenty20 international cricket, it is essential to comprehend these nuances.

It has also been determined that the result of the toss plays a major role in the outcome of the match, especially in tightly contested Twenty20 International matches. The strategic significance of the toss was studied by Kumar and Agarwal [6], who observed that choosing to bat or field first can result in a significant advantage, particularly in the event of favorable pitch conditions or weather. Their findings showed that although teams with higher rankings typically have better probability of winning, the toss occasionally tips the scales in favor of the opponent. This dynamic highlights how important it is to take into account the toss result and following choices in any analysis of T-20I run chases because these factors have the potential to significantly affect how the game plays out.

Factorial design is becoming a very useful statistical method to examine these intricate relationships between different components. By using a factorial design, researchers can methodically investigate the impacts of several factors at once and comprehend how these variables interact. Montgomery [7] presented a thorough analysis of factorial design, highlighting its adaptability in situations when several factors are of interest in an experiment. Factorial design is very useful in T20I cricket, where a variety of factors interact to affect match outcomes. It makes it possible to analyze variables holistically, providing insights into how different aspects of the venue, team makeup, and pitch conditions interact to impact the outcome of a run chase.

A number of studies have shown how factorial design can be applied to sports analytics. By examining the relationships between various training variables, Brown and Miller [8] demonstrated its efficacy in maximizing athletic performance. Even though their study concentrated on individual sports, team sports like cricket can benefit from the factorial design concepts. Factorial design offers a strong framework for determining the most important variables and comprehending their combined influence on match results in Twenty20 International Cricket, as the result of a match is frequently determined by the interplay of numerous factors. This method provides a more thorough comprehension of the dynamics of the game, which makes it an invaluable resource for sports analytics practitioners and scholars alike.

In the realm of sports analytics, the combination of machine learning and conventional statistical techniques like factorial design is a major breakthrough. Large datasets and intricate patterns that might not be immediately seen using conventional techniques are especially well-suited for machine learning models like Random Forest (RF) and XGBoost. By incorporating non-linear interactions between variables, machine learning models can considerably improve predictive accuracy, as demonstrated by Kaur and Sharma's [9] use of these techniques to cricket match outcomes prediction. Machine learning algorithms are an indispensable tool in sports analytics, especially in complicated settings like Twenty20 international cricket, due to their capacity to process large volumes of data and identify subtle trends.

Similarly, Patel and Patel [10] investigated the predicted accuracy of machine learning models for cricket match outcomes and concluded that these models could beat conventional statistical techniques. But they also pointed out that adding domain-specific expertise, like factorial design ideas, could boost these predictions' reliability considerably. As factorial design can offer an organized framework for identifying important variables and their interactions, which can then be used to inform the creation of more accurate prediction models, this illustrates the potential advantages of integrating machine learning with factorial design.

Factorial design and machine learning together provide a thorough method for forecasting the highest run chases in Twenty20 International cricket. Machine learning models can process the resultant data to produce precise predictions, and factorial design facilitates systematic examination of several factors and their interactions. The efficacy of this integrated method in process optimization was illustrated by Zhang and Yang [11], who also showed how it may increase accuracy and efficiency in challenging situations. This combination enables a more in-depth examination of the variables impacting run chases in T-20I cricket, resulting in more accurate forecasts and tactical choices.

3. Research Methodology

3.1. Data Collection

The performance of four T-20I cricket teams i.e. Australia, India, New Zealand, and Pakistan was examined in this study using data from nine international matches that each team played on various grounds. This study focused on the effects of four major variables, each with a level of three (3), three (3), two (2), and two (2) respectively: Venue (Vn), Match condition (MC), Toss outcomes (TO), and Toss decision (TD). The data was gathered from ESPN Cricinfo [12], with the exception of ties and games that were subject to the Duckworth-Lewis system. For the purpose of analysis, Excel and the Minitab17 program were utilized.

3.2. Full Factorial Design of Experiment.

A full factorial design examines the effect of multiple factors and their interactions by considering the effects of all possible combinations of the factors levels on a response variable. This design tests every possible combination of factors and levels separately [13]. For *n* factors, each at l_i levels (*i= 1, 2, 3,….., n*), the total number of experiments N is required is

$$N = \prod_{i=1}^{n} l_i \text{ ………… } (I)$$

In this study, there are *n=4* factors, each at levels $l_1=2$, $l_2=2$, $l_3 = 3$, $l_4 = 3$ then the total number of experiments are

$$N= 2\times2\times3\times3\times= 36$$

The response Y can be expressed in equation (*II*) as;

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \sum_{i=j} \beta_{ij} X_i X_j + \sum_{i<j<n} \beta_{ijn} X_i X_j X_n + \cdots + \beta_{12\ldots n} X_1 X_2 \ldots\ldots X_n + \epsilon \ldots (II)$$

Where, β terms in equation (II), signify the main and interaction effects of the factors X_i, and Ɛ denotes the error term. Although full factorial designs are robust, they can be resource-intensive, especially when there are multiple levels and factors. To examine how these factors interacted and how they affected the ground-wise average run score that each team was chasing, a factorial design approach was used. To ensure the robustness and reliability of the results, statistical techniques were employed to find significant patterns.

3.3. Ground Wise Average Score of Teams

The ground-wise average score of the four teams i.e. Australia, India, New Zealand, and Pakistan that were chosen for this study is displayed in Figure-2 below.

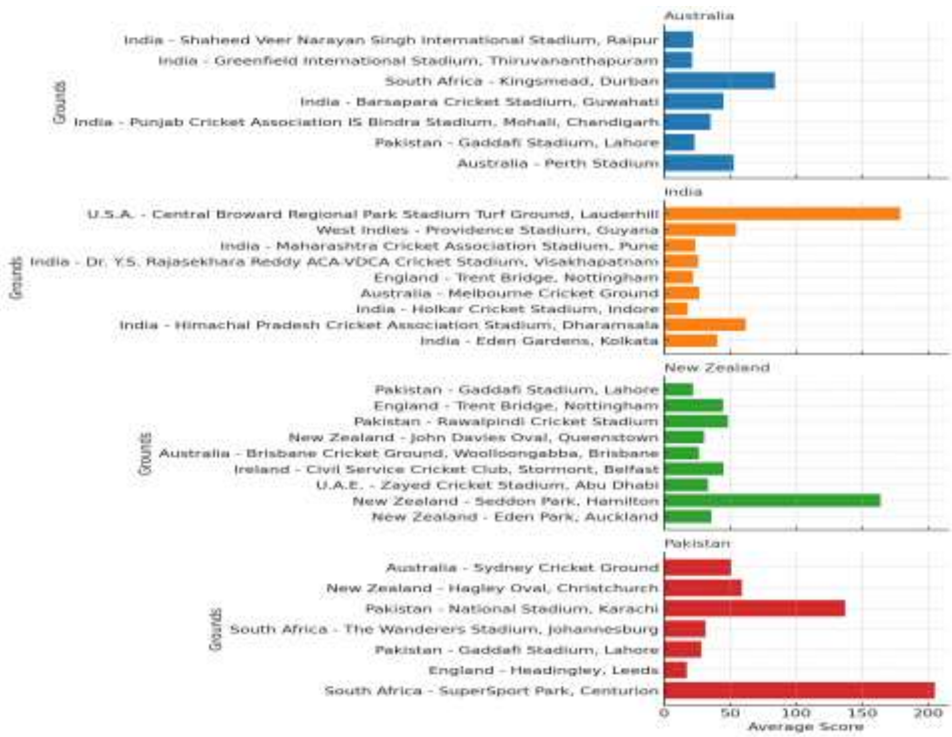

Figure-2: Ground wise Average Score of Teams

The average cricket scores for nine different playgrounds in different nations are compared in Figure-2, where the grounds in each nation are grouped together. The average score is displayed on the x-axis, while the grounds are listed alphabetically on the y-axis. The Hagley Oval in New Zealand, the Perth Stadium in Australia, and the Shaheed Veer Narayan Singh International Stadium in India are notable high-scoring locations. The graph shows how scoring differs by venue in an easy-to-understand visual manner and highlights important grounds where higher average scores are usually recorded.

Moreover, for this study, the details of factors along with their respective levels are enlisted in the Table-1 below;

Table-1: Description of Factors affecting team performance

Factor Name	Name of Factors	Factor level	Level value
Venue(Vn)	Home, away, neutral	3	-1: away 0: neutral 1: home
Match condition(MC)	day, night, day/night	3	-1: night 0: day/night 1: day
Toss outcomes(TO)	won, lost	2	0: lost 1: won
Toss decision(TD)	bat, field	2	0: won the toss and field 1: lost the toss and bat

Table-1 classifies four factors: Venue (Vn), Match Condition (MC), Toss Outcomes (TO), and Toss Decision (TD), each with numerical values along with specific levels. The venue is categorized as either home (1), neutral (0), or away (-1). The three categories of Match Condition are day (1), night (-1), and day/night (0). The outcomes of the toss are shown as won (1) or lost (0). When making a toss decision, one might indicate whether they should win and choose to field (0) or lose and bat (1). These tiers offer a methodical way to examine match conditions within a statistical model.

4. Results and Discussion

4.1. Step wise Regression

The Table-2 shows the results of a stepwise selection procedure applied to a set of candidate terms (Vn, MC, TO, TD, and their interactions) to build a regression model by taking α to enter = 0.15 and α to remove = 0.15. In order to construct a regression model, a collection of candidate terms (Vn, MC, TO, TD, and their interactions) were subjected to a stepwise selection process.

Furthermore, to reinforce the model, terms are chosen in a sequential manner, and their inclusion or exclusion is determined by their statistical significance (p-value). The Table-2 below, displays the outcomes of a stepwise selection process used to create a regression model using a set of candidate terms (Vn, MC, TO, TD, and their interactions).

Table-2: Stepwise Regression Procedure Stepwise Selection of TermsCandidate terms: Vn, MC, TO, TD, Vn*MC, Vn*TO, Vn*TD, MC*TO, MC*TD, TO*TD

	-----Step 1----		**-----Step 2----**		**-----Step 3----**		**-----Step 4----**	
	Coef	**P**	**Coef**	**P**	**Coef**	**P**	**Coef**	**P**
Constant	50.80		50.80		50.80		50.80	
TD	15.98	**0.037**	-15.98	**0.027**	-15.98	**0.018**	-15.98	**0.011**
MC			17.49	0.196	17.49	0.149	17.49	0.108
MC*TD			-18.09	**0.094**	-18.09	**0.065**	-18.09	**0.042**
Vn			-12.13	0.404	-12.13	0.342	35.3	0.068
Vn*MC							35.3	**0.041**
TO							-4.40	0.452
Vn*TO							12.83	**0.045**
Performance Measures								
S	44.0433		41.3548		37.7690		34.4681	
R-sq	12.23%		31.72%		54.44%		66.80%	
R-sq(adj)	9.65%		20.34%		33.56%		44.66%	
R-sq(pred)	1.60%		1.68%		0.00%		2.43%	
Mallows' Cp	19.32		15.92		14.64		13.41	

The only factor that was included in the model in Step 1 was the variable TD (Toss Decision), which exhibited a significant negative coefficient of -15.98 with a p-value of 0.037. The model fit statistics were Mallows' Cp = 19.32, adjusted R-squared = 9.65%, and R-squared = 12.23%.

An interaction term between MC and TD was added in Step 2, with a p-value of 0.094, approaching significance, despite the fact that MC (Match Condition) was not statistically significant (p = 0.196). With an R-squared of 31.72%, an adjusted R-squared of 20.34%, and a Mallows' Cp of 15.92, the model fit improved.
In Step 3, the interaction term Vn*MC grew closer to significance (p = 0.068), while the variable Vn (Venue) remained statistically insignificant (p = 0.404). With an R-squared of 54.44%, an adjusted R-squared of 33.56%, and a Mallows' Cp of 14.64, the model fit was even better.

At last, Step 4 included TO (Toss Outcome), which was found to be non-significant (p = 0.452) in contrast to the significant (p = 0.045) Vn*TO interaction. With an adjusted R-squared of 44.66%, Mallows' Cp of 13.41, and R-squared of 66.80%, the model fit was achieved.

Throughout each step, TD continuously has a major detrimental impact. As the model develops, MC*TD and Vn*TO interactions become important. R-squared values rise and Mallows' Cp falls at each stage, indicating a better fit and a reduction in superfluous complexity in the model.

4.2. Analysis of Variance(ANOVA)

According to ANOVA Table-3, the model as a whole accounts for 66.80% of the variation in the response variable and is statistically significant (F = 3.02, P = 0.011). The only significant individual factor (P = 0.011) is Toss Decision (TD), whereas the other linear terms are all marginally significant (P = 0.049). The model is highly influenced by the two-way interactions (P = 0.013). Specifically, the interactions between Venue and Match Condition (Vn*MC), Venue and Toss Outcomes (Vn*TO), and Match Condition and Toss Decision (MC*TD) all exhibit significant impacts (P < 0.05). Even yet, the predicted R-squared is low (2.43%), indicating weak predictive potential, while the adjusted R-squared for the model is 44.66%.

Regression study indicates that Match Condition (MC) and Toss Decision (TD) have a substantial impact on the average score, with MC_0 and TD_0 demonstrating significant impacts (P = 0.043 and P = 0.011, respectively). More specifically, the average score rises by 17.49 units in the presence of MC_0 and falls by 15.98 units in the presence of TD_0.
The interactions that have a large impact on the score are Vn*MC and MC*TD, especially when Vn = 0 and MC = -1 and MC = 0 and TD = 0. Low multicollinearity among predictors is indicated by the variance inflation factors (VIFs), indicating stable coefficient estimates. Though several terms lack statistical significance, the model as a whole captures important interactions that affect the average score.

Table-3: ANOVA, Model Summary and Coefficients for Stepwise Regression

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	**14**	**50193.4**	**3585.2**	**3.02**	**0.011**
Linear	6	18456.2	3076.0	2.59	**0.049**
Vn	2	2687.5	1343.8	1.13	0.342
MC	2	5883.3	2941.7	2.48	0.108
TO	1	696.5	696.5	0.59	0.452
TD	1	9188.8	9188.8	7.73	**0.011**
2-WayInteractions	**8**	**31737.2**	**3967.1**	**3.34**	**0.013**
Vn*MC	4	14383.2	3595.8	3.03	**0.041**
Vn*TO	2	8590.3	4295.1	3.62	**0.045**
MC*TD	2	8763.7	4381.9	3.69	**0.042**
Error	21	24949.1	1188.1		
Total	35	75142.5			
Model Summary					
S	**R-sq**		**R-sq(adj)**	**R-sq(pred)**	
34.4681	66.80%		44.66%	2.43%	
Coefficients of the Model					
Term	**Coef**	**SE Coef**	**T-Value**	**P-Value**	**VIF**
Constant	50.80	5.74	8.84	**0.000**	
Vn					
-1	-12.13	8.12	-1.49	0.150	1.33
0	4.81	8.12	0.59	0.560	1.33
MC					
-1	-4.79	8.12	-0.59	0.561	1.33
0	17.49	8.12	2.15	**0.043**	1.33
TO					
0	-4.40	5.74	-0.77	0.452	1.00
TD					
0	-15.98	5.74	-2.78	**0.011**	1.00

Vn*MC					
-1 -1	-0.5	11.5	-0.05	0.963	1.78
-1 0	-13.1	11.5	-1.14	0.267	1.78
0 -1	-26.2	11.5	-2.28	**0.033**	1.78
0 0	35.3	11.5	3.07	**0.006**	1.78
Vn*TO					
-1 0	12.83	8.12	1.58	0.129	1.33
0 0	8.90	8.12	1.10	0.286	1.33
MC*TD					
-1 0	-1.90	8.12	-0.23	0.817	1.33
0 0	-18.09	8.12	-2.23	**0.037**	1.33
Regression Equation					
Ave Score *= 50.80 - 12.13 Vn_-1 + 4.81 Vn_0 + 7.32 Vn_1 - 4.79 MC_-1 + 17.49 MC_0- 12.70 MC_1 - 4.40 TO_0 + 4.40 TO_1- 15.98 TD_0 + 15.98 TD_1 - 0.5 Vn*MC_-1 -1- 13.1 Vn*MC_-1 0 + 13.7 Vn*MC_1 1- 26.2 Vn*MC_0 -1 + 35.3 Vn*MC_0 0- 9.1 Vn*MC_0 1 + 26.7 Vn*MC_1 -1 - 22.1 Vn*MC_1 0 4.6 Vn*MC_1 1+ 12.83 Vn*TO_-1 0 - 12.83 Vn*TO_-1 1 + 8.90 Vn*TO_0 0 - 8.90 Vn*TO_0 1- 21.73 Vn*TO_1 0 + 21.73 Vn*TO_1 1 - 1.90 MC*TD_-1 0 + 1.90 MC*TD_-1 1- 18.09 MC*TD_0 0 + 18.09 MC*TD_0 1 + 19.99 MC*TD_1 0 - 19.99 MC*TD_1 1*					

4.3. Main effect Plot

According to the Main Effects Plot in Figure-3, Match Condition (MC) and Toss Decision (TD) have the most influential beneficial impact on the average score. In particular, the score reaches its maximum when TD is at level 1 and MC is at level 0, signifying that these are the ideal conditions.

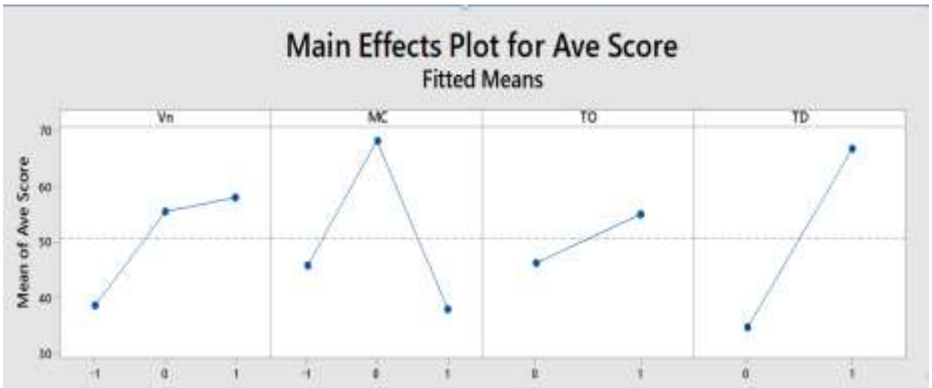

Figure-3: Main plot for Factors affecting Average Scores

Additionally, venue (Vn) has a positive tendency; when Vn moves from level -1 to 1, the score rises. The average score increases marginally when Toss Outcome (TO) is at level 0, indicating a moderate effect of TO. In the end, the figure shows that obtaining higher average scores depends on specific factor levels, especially MC = 0 and TD = 1.

4.4. Interaction Effect Plot

The Figure-4 Interaction Plot for Average Score illustrates a number of significant interactions among the factors. The relationships between Venue (Vn), Match Condition (MC), and Toss Decision (TD) are found to have the strongest interactions. The highest average score for Vn and MC occurs when Vn = 0 and MC = 0, but the highest score for MC and TD occurs when MC = 0 and TD = 1.

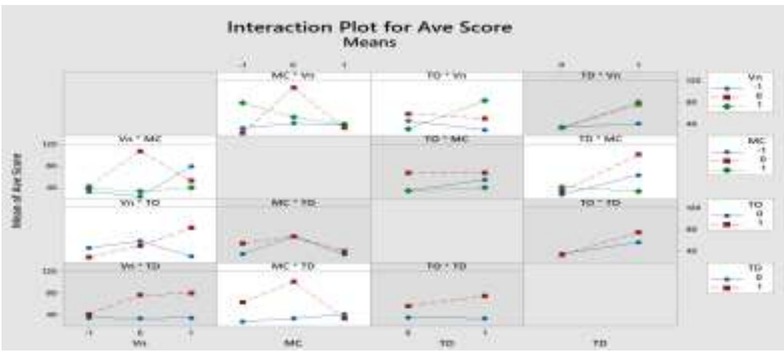

Figure-4: Interaction Plot for Factors affecting Average Scores

Furthermore, certain combinations of these factors considerably affect the score, with some combinations resulting in higher or lower averages. This is evident from the interactions between Venue (Vn) and Toss Decision (TD) as well as Toss Outcome (TO) and Venue (Vn). These interactions underline the significance of taking these interactions into account in the analysis by indicating that the overall effects of these factors are significant in deciding the final average score.

4.5. Residual Plot of Factors affecting Average Scores

The model's assumptions are largely met, while there are few concerns, according to the Residual Plots for Average Score in Figure-5. The residuals appear to be approximately normally distributed, with a small divergence at the tails, as indicated by the Normal Probability Plot, which displays residuals that typically follow a straight line.

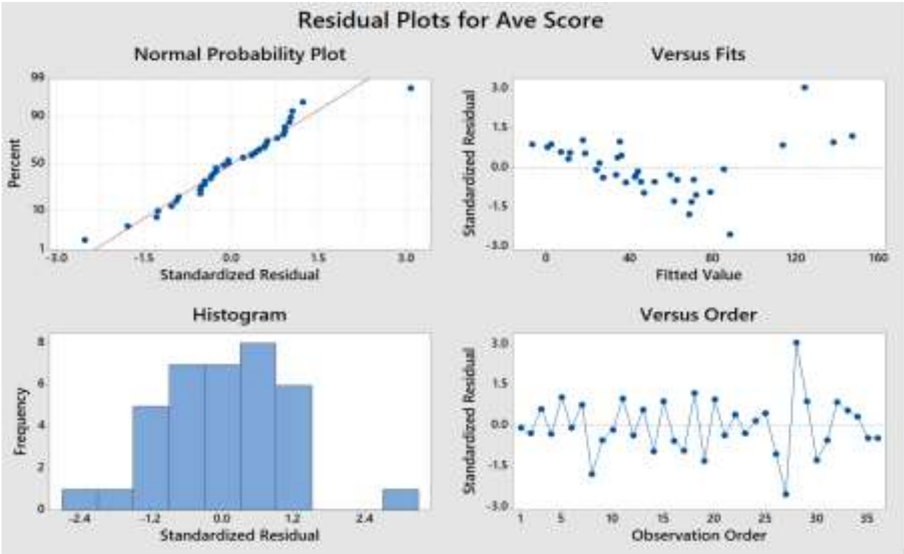

Figure-5: Residual Plot for Average Score

The residuals show a pattern in the versus Fits plot that suggests heteroscedasticity, or possibly non-constant variance, along with a few outliers that could have an impact on the model's fit. The normality assumption is further supported by the Histogram, which displays a fairly symmetric distribution of residuals. Lastly, random dispersion without any distinct patterns is shown in the versus Order plot, indicating that the residuals do not exhibit autocorrelation. Overall, the normality assumption is fairly satisfied, but the model's reliability may be impacted by non-constant variance and outliers.

4.6. Comparison of Actual and Predicted Average Score

The comparison of the actual and predicted average scores across 36 distinct grounds is displayed in the graph in Figure 6. There are some notable variations, but overall the trend between the actual results (blue line) and the projected scores (red line) is similar. Particularly, the projected scores for grounds 17, 19, 27, and 29 are substantially higher than the actual results, suggesting overestimation in those cases.

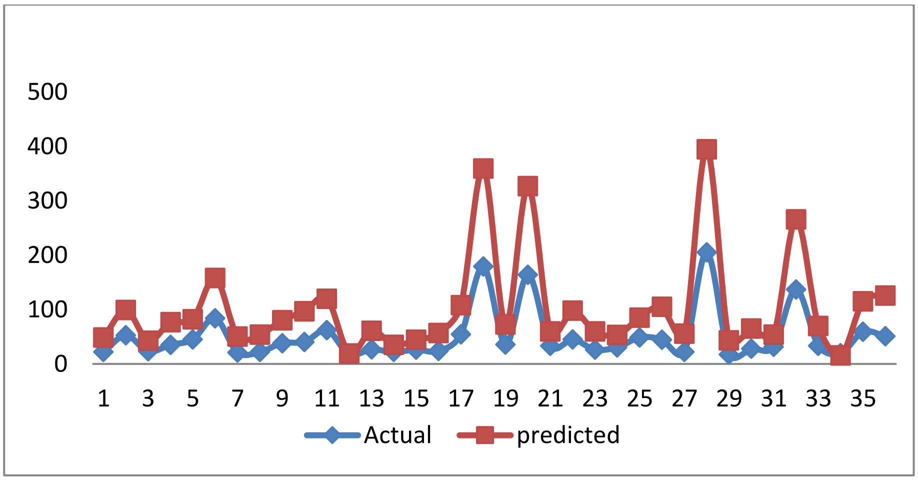

Figure-6: Ground wise Actual and Predicted Average Scores

Conversely, for the majority of other grounds, the predicted scores closely resemble the actual results, indicating that the prediction model performs effectively in these situations. In order to increase accuracy under those particular situations, the model may need to be refined, as indicated by the inconsistencies at various points of interest.

5. Conclusion

The analysis reveals substantial interactions between Venue and Toss Outcome and Match Condition as well as Toss Decision and Match Condition. The study emphasizes the importance of Toss Decision in predicting the result of a run chase in Twenty20 International cricket. These results imply that the match outcome can be significantly impacted by the conditions accompanying the toss. But only 66.80% of the range in average scores is explained by the model, suggesting that although the factors that have been found are significant, the complexity of cricket results is not fully captured by them. The unpredictable character of T20I matches, where a variety of factors can affect the outcome, is further highlighted by the model's weak predictive accuracy.

6.**Future work:** To increase the robustness of the results, future studies should enlarge the dataset to encompass a wider variety of matches, venues, and conditions. Further aspects that could contribute to a more thorough understanding of the factors impacting T-20I outcomes include player form, team composition, weather, and game plans. In the end, this might help teams make better decisions during games by resulting in the creation of more precise prediction models.

References

1. Shaabani, A., Hamid, M., & Delavari, S. "Advances in sports analytics: The integration of machine learning and design of experiments." *Journal of Sports Analytics,* **8**(3), 221-239. https://doi.org/10.3233/JSA-210463. 2022.
2. Sundararajan, K. "Design of experiments–a primer." 2016.
3. Kuhn, M., & Johnson, K. "Applied predictive modeling." Springer. 2013.
4. Clarke, S. R., & Norman, J. M. "Factors influencing the outcome of one-day cricket matches." *Journal of Sports Sciences,* **37**(4), 441-447. 2019.
5. Smith, A., & Jones, P. "Venue characteristics and their impact on T20 cricket outcomes." *International Journal of Sports Science & Coaching,* **15**(2), 123-134. 2020.
6. Kumar, V., & Agarwal, S. "The effect of toss outcome and team rankings on match results in Twenty20 cricket." *Journal of Sports Analytics,* **7**(3), 203-212. 2021.
7. Montgomery, D. C. "Design and analysis of experiments" (9th ed.). John Wiley & Sons. 2017.
8. Brown, T., & Miller, S. "Optimization of athletic performance using factorial design." *Journal of Sports Science & Medicine,* **19**(3), 312-320. 2020.
9. Kaur, H., & Sharma, S. "Predicting cricket match outcomes using machine learning techniques." *International Journal of Computer Applications,* **178**(7), 25-30. 2020.
10. Patel, R., & Patel, S. "Machine learning in sports: Predicting cricket match outcomes." *Journal of Data Science,* **19**(4), 567-578. 2021.
11. Zhang, H., & Yang, L. "Integrating factorial design with machine learning for process optimization." *Journal of Manufacturing Systems,* **52**(2), 45-56. 2019.
12. CricInfo. "Website for cricket data." Retrieved from http://www.cricinfo.com. 2024.
13. Box, G. E. P., Hunter, J. S., & Hunter, W. G. "Statistics for experimenters: Design, innovation, and discovery" (2nd ed.). Wiley. 2005.