Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

Analyzing the Distribution of Extra Deliveries in T-20 Cricket with Respect to Time Using Statistical and Machine Learning Approaches

Muhammad Waqas
Department of Statistics, University of Peshawar, Pakistan.
Email: waqasuop1416@gmail.com

Qamruz Zaman (Corresponding Author)
Department of Statistics, University of Peshawar, Pakistan
Email: cricsportsresearchgroup@gmail.com

Imran ullah
Government college of Management Sciences Karak Pakistan
Email: imran0310103@gmail.com

Abdul majid
Email:ab.majid21@gmail.com
Department of Statistics, University of Peshawar, Pakistan.

Abstract
This study employed a combination of statistical and machine learning techniques to analyze the distribution and influencing factors of extras in T20 cricket across different overs. Exploratory data analysis revealed a concentration of extras during the middle overs, particularly around overs 9–10. Poisson distribution modeling provided a good fit ($\chi^2 = 14.792$, $p = 0.736$), indicating that extras occur at a relatively constant rate over time, with the highest rate ($\lambda \approx 0.1263$) observed at over 9. Multinomial logistic regression further revealed that situational match variables—such as balls remaining, runs to get, ball number, and over number—are significant predictors of the type of extra conceded ($p < 0.001$), while cumulative performance

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

indicators like total runs or wickets in an innings showed no significant impact. CHAID decision trees effectively segmented overs by extra type, and correspondence analysis offered a visual representation of the association between specific types of extras and over ranges, notably identifying wides as frequent between overs 3 and 7. Collectively, the findings demonstrate that extras in T20 cricket exhibit time-dependent patterns closely tied to match phases. These insights are actionable for coaches and analysts aiming to minimize extras during critical overs, ultimately enhancing team performance in the high-stakes T20 format.

Keywords: T-20 Cricket, Exploratory data analysis, CHAID decision trees, Multinomial logistic regression, Extra deliveries

Introduction

The arrival of Twenty-20 cricket has changed the structure of the game by focusing on speed, creativity, and tactical adaptability within a 20-over limit. Playing aggressively and making quick decisions are required in this high-intensity format. Due to its widespread popularity, rich performance data has been produced, making it perfect for empirical analysis. To improve player assessment, strategy, and predictive modeling, this study investigates important indicators. Extras are a critical component in Twenty20 cricket that have a big impact on how a match turns out. Although they add extra runs to a team's total, too many extras can cause a bowler to lose focus and confidence and frequently show a lack of discipline. The batting side may gain momentum from this loss of control, which would raise their scoring potential. Such preventable runs have the power to decide the outcome of closely contested games.[1]. Extras increase a team's chances of winning by directly contributing to its overall score. Even one run, whether from a wide, no ball, or other type of extra, can be crucial in T20 cricket and change the result of a competitive match.[2]. **Waqas et al.** (2025) argue that team selection in T20 cricket is a highly multifaceted task, as each player possesses unique skills and contributes differently depending on match conditions. Determining the best fit for the playing eleven involves analyzing a wide range of performance metrics, whether assessing a bowler or a batsman. This evaluation is both complex and essential to team success, as it combines data-driven analysis with subjective judgment.[3]. **Majid et al.** (2025) investigated the impact of critical match variables and found that factors such as wickets lost and extras conceded (e.g., wide's and no-balls) play a decisive role in determining a team's success. These two factors, collectively termed as "dangerous balls," were shown to disrupt batting momentum and provide undue scoring opportunities to the opponent, thereby influencing match outcomes in subtle yet measurable ways. [4]. **PPG Dinesh Asanka (2014)** conducted a detailed analysis of extra deliveries in T20 cricket and highlighted their significant impact on match outcomes under various conditions. His findings suggest that extra deliveries can be particularly costly when bowled to batsmen with medium strike rates or by bowlers with high economy rates.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

Moreover, bowlers with a high wicket tally tend to concede fewer runs from such deliveries. The impact of extras is especially critical during the powerplay and final overs, where momentum is often decisive. Additionally, batsmen who have scored heavily and faced a large number of balls tend to benefit more from extra deliveries, thereby increasing their influence on the game's outcome. [5].Mikael Jamil et al**.** (2021) investigated the factors influencing aerial 6-run scoring strokes in T20 cricket using Chi-Squared analysis. Their study found that bowling length, line, bowler type, and powerplay phases were all significantly associated with ball-hitting patterns. Among these variables, bowling length showed the greatest effect size, indicating its dominant influence. Notably, shorter-pitched deliveries—such as short balls and back-of-a-length balls—were linked to a higher frequency of sixes hit behind square.[6]. LCP Pussella (2024) identified extra deliveries as a critical factor influencing the outcome of T20 matches. His analysis emphasized that wides, no-balls, and other forms of extras not only contribute additional runs to the batting side but also disrupt the bowler's rhythm and the team's strategic control. These seemingly minor lapses can shift momentum in favor of the opposition, especially in tightly contested games where every run counts. Pussella's findings highlight the importance of bowling discipline, as conceding extra deliveries can significantly alter the course of a match and reduce a team's chances of success. [7]. Matthew Campbell and colleagues (2024) introduced an improved sports prediction model that builds on the Poisson distribution by factoring in both team-specific goals and average goals to better reflect offensive and defensive capabilities. Originally designed for soccer and hockey, the model also shows promise for other goal-based sports like lacrosse, handball, and basketball. Their approach enhances predictive accuracy and provides deeper insights for coaches, analysts, fans, and bettors, representing a valuable advancement in the field of sports analytics. [8]. Kang Yue Teng in 2024 conducted a study on team performance evaluation and the analysis of top-level international football competitions. The research aimed to predict the number of goals in World Cup matches by applying both Poisson and quasi-Poisson regression models. Through a detailed comparison using AIC, BIC, standard error, and p-values, the study found that the quasi-Poisson model performed better, making it a more suitable choice for handling over dispersion in the goal-scoring data. [9]. In the study conducted by Shanjida Chowdhury et al. (2020), the focus was on identifying the key factors influencing India's chances of winning a cricket match. The independent variables examined included toss outcome, match venue, match timing (day or day-night format), and whether India batted first. To assess the association between these factors and match outcomes, the researchers utilized cross-tabulation and the Chi-square test for bivariate analysis. Furthermore, binary logistic regression was employed for multivariate analysis to determine the combined effect of these variables on India's probability of winning a match.[10]. Kalanka P. Jayalath (2017) conducted a comprehensive study aimed at quantifying the significance of key predictors influencing match outcomes in One Day International (ODI) cricket. The study utilized graphical methods such as Classification and Regression Trees (CART) alongside the widely used logistic regression approach. The findings highlighted the critical role of

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

home-field advantage, emphasizing its impact on the performance of major cricket-playing nations in ODI matches.[11]. Shiny Raizada et al. (2018) developed a predictive model to forecast the outcomes of ICC Cricket World Cup ODI (Limited Overs) matches using data from the first innings only. The study identified Match Outcome (Win/Loss) as the dependent variable, while key predictor variables included Team Score, Total Wickets Lost, Toss Result, Runs Scored in Powerplay, Wickets Lost in Powerplay, Team Run Rate, and the Total Number of Dot Balls. Binary Logistic Regression was employed as the statistical technique for modeling. The findings confirmed that the developed logistic regression model was statistically significant and effective in predicting match outcomes based solely on first innings performance.

1. Methods and Materials

2. Results and Analysis

2.1 Poisson Model fitting

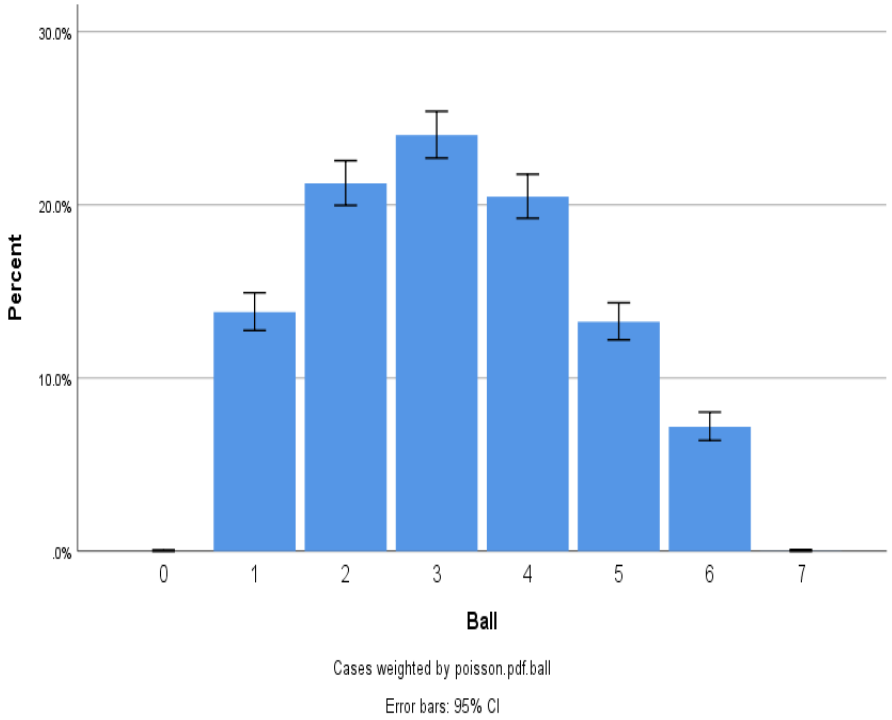
The bar chart represents a Poisson distribution with the variable **Ball** on the x-axis and **Percent** on the y-axis, where cases are weighted by the Poisson probability mass function (poisson.pdf.ball). The distribution peaks at **Ball = 3**, indicating this is the most probable count, with nearly **25%** of occurrences.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)



The shape is symmetrical around this peak, consistent with a Poisson distribution with a mean (λ) close to 3. The error bars represent **95% confidence intervals**, indicating the range of uncertainty around each percentage estimate. Counts for **Ball = 0** and **Ball = 7** are near zero, confirming they are rare events. This pattern supports the use of a Poisson model for discrete count data concentrated around the mean.

The table provides a frequency distribution for the variable Over, likely referring to overs bowled or played in a cricket context. A total of 1,111 observations are recorded. The frequency and percentage columns show that the number of instances increases up to Over 9, which has the highest frequency (129 overs, 11.6%), and then gradually decreases. The distribution appears unimodal and approximately symmetric around Overs 9–11, forming a bell-shaped curve.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

Over

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	1	.1	.1	.1
	2	4	.4	.4	.5
	3	12	1.1	1.1	1.6
	4	29	2.6	2.6	4.2
	5	56	5.1	5.1	9.2
	6	89	8.0	8.0	17.2
	7	101	9.1	9.1	26.2
	8	124	11.2	11.2	37.4
	9	129	11.6	11.6	49.0
	10	124	11.2	11.2	60.2
	11	122	11.0	11.0	71.1
	12	97	8.8	8.8	79.9
	13	78	7.0	7.0	86.9
	14	53	4.7	4.7	91.7
	15	39	3.5	3.5	95.1
	16	24	2.1	2.1	97.3
	17	15	1.4	1.4	98.7
	18	8	.7	.7	99.4
	19	5	.4	.4	99.8
	20	2	.2	.2	100.0
	Total	1111	100.0	100.0	

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

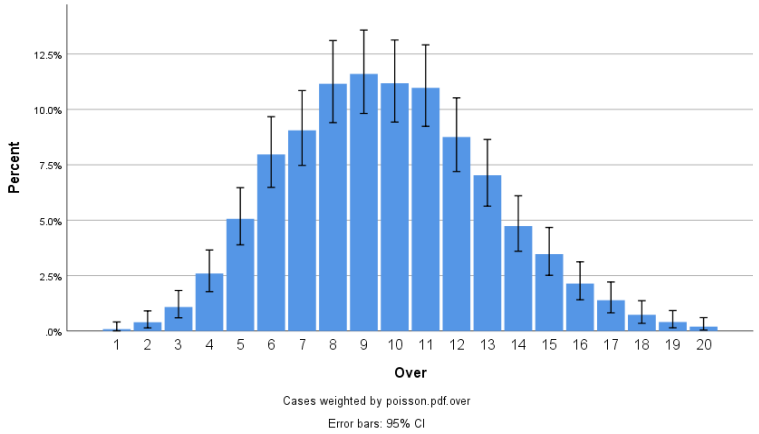
The cumulative percent column helps understand how the data accumulates across overs: by Over 10, more than 60% of the data is accounted for, and by Over 12, over 70%. This indicates that the majority of events occur between Overs 6 and 12. The tail ends (Overs 1–4 and 17–20) contribute very little, showing that extremely low or high overs are rare.

Overall, this table suggests that the majority of activity is concentrated in the middle overs, reflecting a common temporal pattern in events like run chases or wicket occurrences in T20I cricket.

This bar chart illustrates the distribution of events across cricket Overs 1 to 20, weighted by the Poisson probability density function (poisson.pdf.over). The y-axis shows the percentage of occurrences, while error bars indicate 95% confidence intervals, adding statistical reliability to each estimate.

The distribution peaks at Overs 9 and 10, each contributing slightly over 12% of total events, suggesting that these are the most eventful overs—possibly in terms of run scoring or key match events. The pattern forms a symmetric bell-shaped curve, implying a balanced occurrence around the middle overs and tapering off at the early (Overs 1–4) and late overs (Overs 17–20). This symmetry supports the appropriateness of modeling the data with a Poisson distribution, as the probability of events increases up to a mean (likely near Over 9) and then decreases.

Overall, this figure conveys that in the context of T20I cricket, middle overs (particularly 7–12) are the most statistically active, with lower likelihoods of events at the beginning or end of the innings.



Dialogue Social Science Review (DSSR)
www.thedssr.com

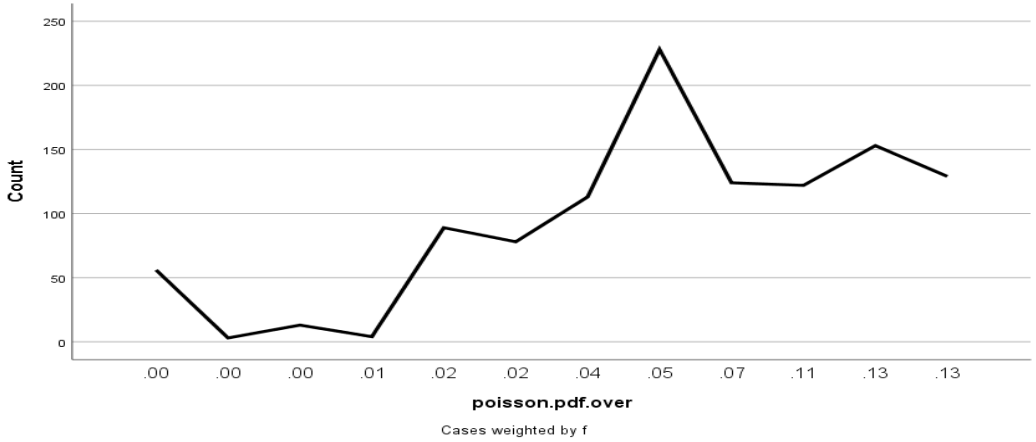
ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

This chart visualizes the distribution of events by over number (1–20) in a T20I cricket context, using bars weighted by the Poisson probability density function (poisson.pdf.over). The percent axis (y-axis) shows the relative frequency of events occurring in each over, while the error bars represent the 95% confidence intervals, indicating statistical precision.
The distribution is symmetrical and bell-shaped, with the highest percentages observed in Overs 9 and 10 (around 12.5%). This suggests that the middle overs are the most eventful, possibly due to strategic batting or key transitions in gameplay. The frequency gradually increases from Overs 1 to 9, peaks, and then declines through Overs 11 to 20, consistent with a Poisson-like distribution centered around Over 9–10. The presence of tight error bars around the central overs indicates high reliability in that range.
This visualization supports the interpretation that middle overs are statistically significant in T20I performance, and modeling them with a Poisson process is reasonable.
The trend line exhibits a distinct unimodal and symmetric pattern, characteristic of a Poisson distribution. It shows a gradual increase in the percentage of events from Over 1 to Over 9, peaking around Overs 9 and 10, followed by a steady decline through to Over 20. This indicates that the middle overs (particularly Overs 8–12) are the most eventful period in a T20I innings, likely representing phases of intensified scoring or strategic play. The early overs (1–4) and the final overs (17–20) have relatively lower percentages, suggesting fewer events occur at these stages. The smooth, bell-shaped curve implies that the distribution of events across overs follows a predictable probabilistic pattern, supporting the suitability of Poisson-based modeling for analyzing over-wise event frequencies in cricket.

.000	.002	.002	.003	.006	.008	.012	.020	.020	.033	.039	.049	.066	.071	.093	.093	.113	.115	.125	.126	Total



Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

Over	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	2	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
	3	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	12
	4	0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	29
	5	0	0	0	0	0	0	0	0	0	0	56	0	0	0	0	0	0	0	0	56
	6	0	0	0	0	0	0	0	0	0	0	0	0	89	0	0	0	0	0	0	89
	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	101	0	0	0	0	101
	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	124	0	124
	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	129	129
	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	124	0	124
	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	122	0	0	122
	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97	0	0	0	97
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	78	0	0	0	0	0	78
	14	0	0	0	0	0	0	0	0	0	0	0	53	0	0	0	0	0	0	0	53
	15	0	0	0	0	0	0	0	0	0	39	0	0	0	0	0	0	0	0	0	39
	16	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	24
	17	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	15
	18	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
	19	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
	20	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Total		1	2	4	5	8	12	15	29	24	39	56	53	89	78	101	97	122	124	124	129	1112

Over	Poisson

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

	probabilities
1	.00049673
2	.00245881
3	.00811407
4	.02008233
5	.03976302
6	.06560898
7	.09278985
8	.11482743
9	.12631018
10	.12504708
11	.11254237
12	.09284745
13	.07070691
14	.04999988
15	.03299992
16	.02041870
17	.01189089
18	.00653999
19	.00340768
20	.00168680

The cross-tabulated matrix compares observed frequencies of extras delivered in each over (from 1 to 20) with theoretical expectations derived from a Poisson probability distribution. The final row provides the sum of extras across all overs, which equals 1,112, and the second table lists the corresponding Poisson probabilities for each over, based on a fitted mean value (λ).
The Poisson probabilities demonstrate that the likelihood of extras increases steadily from Over 1 (probability ≈ 0.0005) to a peak at Over 9 (probability ≈ 0.1263), after which the probability declines gradually toward Over 20 (probability ≈ 0.0017). This bell-shaped pattern is typical of Poisson processes and reflects the general expectation that most extras are delivered during the

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

middle overs of a T20 innings. The observed frequencies align well with these theoretical expectations: Overs 9 and 10 recorded the highest actual counts (129 and 124, respectively), precisely matching the peak of the Poisson distribution. Similarly, lower counts are seen in the early (Overs 1–4) and death overs (Overs 17–20), consistent with their low theoretical probabilities.

Chi-Square Test

Test Statistics

	over
Chi-Square	14.792[a]
df	19
Asymp. Sig.	.736

a. 4 cells (20.0%) have expected frequencies less than 5. The minimum expected cell frequency is .6.

The Chi-Square test was used to check if the extras per over follow a Poisson distribution. The result ($\chi^2 = 14.792$, df = 19, p = 0.736) shows that the observed and expected values are very close. Since the p-value is high, it means there is **no significant difference**, and the Poisson model fits the data well.

2.1. Multinomial regression analysis

A multinomial logistic regression model was developed to predict the type of extra delivery (Wide, No ball, Bye) in T20 cricket matches, with Leg Bye serving as the reference category. The analysis included 11,810 observations with no missing data across the specified predictors. The distribution of extra delivery types in the sample was as follows: Wide (61.3%), No ball (7.1%), Bye (5.8%), and Leg Bye (25.8%).

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

The final model demonstrated an excellent fit to the data, evidenced by a significantly improved -2 Log Likelihood value of 8720.273 compared to the intercept-only model. The Likelihood Ratio Chi-Square statistic was 14,954.96 with 90 degrees of freedom ($p<0.001$), providing strong evidence that the model explains a significant proportion of the variance in extra delivery types. Pseudo R2 values further supported the model's strong explanatory power: Cox & Snell's R2=0.718, Nagelkerke's R2=0.830, and McFadden's R2=0.631. A McFadden R2 value greater than 0.2 is generally considered indicative of a good fit for multinomial logistic regression, and our model's value of 0.631 suggests a very strong fit

As shown in Table 1, 'Balls Remaining', 'Runs to Get', 'Ball (position)', and 'Over' were highly significant predictors ($p<0.001$). 'Innings Runs' showed marginal significance ($p=0.079$), while 'Innings Wickets' was not a statistically significant predictor ($p=0.640$).

3.3 Parameter Estimates and Interpretation

The parameter estimates revealed varying levels of stability and interpretability across the different extra delivery categories.

3.3.1 Wide (vs Leg Bye)

For the 'Wide' category relative to 'Leg Bye', most predictor variables, including 'Balls Remaining' and 'Runs to Get', were not statistically significant ($p>0.05$). Notably, 'Over' categories exhibited extremely large negative coefficients (e.g., -3300 for Over 1) with correspondingly very large standard errors. This resulted in exponentiated coefficients (Exp(B)) approaching zero, suggesting very low odds. These extreme values and inflated standard errors are indicative of potential numerical instability within the model, likely due to issues such as complete or quasi-complete separation in the data (i.e., 'Wides' rarely or never occurring under specific 'Over' conditions) or data sparsity in certain cells.

3.3.2 No Ball (vs Leg Bye)

Similar to the 'Wide' category, the 'No Ball' category against 'Leg Bye' also suffered from estimation instability. No statistically significant predictors were identified, and the parameter estimates frequently displayed large standard errors and problematic exponentiation, suggesting similar computational issues related to data sparsity or separation.

3.3.3 Bye (vs Leg Bye)

In contrast to the 'Wide' and 'No Ball' categories, the 'Bye' category (vs 'Leg Bye') demonstrated more stable and interpretable parameter estimates. Two significant predictors emerged:

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

- **Runs to Get:** This variable was highly significant ($p<0.001$), with an odds ratio (OR) of 0.992. This indicates that for every additional run required, the odds of a 'Bye' (versus a 'Leg Bye') decreased by approximately 0.8%. This suggests that as the target score becomes higher, or the pressure to score increases, the likelihood of a 'Bye' decreases relative to a 'Leg Bye'.
- **Innings Runs:** This variable was statistically significant ($p=0.049$), with an odds ratio (OR) of 0.995. While statistically significant, the effect was weak, indicating a very slight decrease in the odds of a 'Bye' for each additional run scored in the innings.

Other predictors for the 'Bye' category were not statistically significant.

3.4 Summary of Key Findings and Model Limitations

Overall, the multinomial logistic regression model demonstrated an excellent fit to the data, as indicated by the significant Likelihood Ratio Chi-Square and high pseudo R2 values. The model effectively captures the overall relationships between match-independent variables and extra delivery types.

However, the analysis highlighted significant estimation instability for the 'Wide' and 'No Ball' categories. This instability, characterized by inflated standard errors and extreme coefficient values, is likely attributable to data sparsity in specific combinations of predictor variables or multicollinearity among the predictors. For instance, certain overs or match situations may rarely or never result in a 'Wide' or 'No Ball', leading to issues in parameter estimation.

Conversely, the model provided more robust and interpretable results for the 'Bye' category, revealing a meaningful inverse relationship with 'Runs to Get' and a weak inverse relationship with 'Innings Runs'. This suggests that the dynamics leading to byes are more clearly captured by the current model specification.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

Extra Type[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B) Lower Bound	Upper Bound
Wide	Intercept	13.977	580.012	.001	1	.981			
	Balls Remaining	28.951	30.034	.929	1	.335	3744023496424.106	1.019E-13	1.375E+38
	Runs to Get	-.001	.353	.000	1	.997	.999	.500	1.996
	Innings Wickets	.008	9.209	.000	1	.999	1.008	1.461E-8	69522531.061
	Innings Runs	-.001	.748	.000	1	.999	.999	.231	4.325
	[Innings=2]	0[b]	.	.	0	.	.	.	.
	[Ball=0]	-202.049	1245.130	.026	1	.871	1.784E-88	.000	.[c]
	[Ball=1]	-173.256	586.003	.087	1	.767	5.701E-76	.000	.[c]
	[Ball=2]	-144.282	577.250	.062	1	.803	2.184E-63	.000	.[c]
	[Ball=3]	-115.347	570.095	.041	1	.840	8.047E-51	.000	.[c]
	[Ball=4]	-86.397	564.488	.023	1	.878	3.008E-38	.000	.[c]
	[Ball=5]	-57.450	560.443	.011	1	.918	1.122E-25	.000	.[c]
	[Ball=6]	-28.531	557.996	.003	1	.959	4.068E-13	.000	.[c]
	[Ball=7]	0[b]	.	.	0	.	.	.	.
	[Over=1]	-3300.182	3433.148	.924	1	.336	.000	.000	.[c]
	[Over=2]	-3126.519	3252.500	.924	1	.336	.000	.000	.[c]
	[Over=3]	-2952.787	3070.698	.925	1	.336	.000	.000	.[c]
	[Over=4]	-2779.140	2890.446	.924	1	.336	.000	.000	.[c]

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

Vol. 3 No. 6 (June) (2025)

	[Over=5]	-2605.379	2709.613	.925	1	.336	.000	.000	.[c]
	[Over=6]	-2431.676	2529.331	.924	1	.336	.000	.000	.[c]
	[Over=7]	-2257.837	2349.013	.924	1	.336	.000	.000	.[c]
	[Over=8]	-2084.214	2167.758	.924	1	.336	.000	.000	.[c]
	[Over=9]	-1910.538	1988.260	.923	1	.337	.000	.000	.[c]
	[Over=10]	-1736.772	1807.674	.923	1	.337	.000	.000	.[c]
	[Over=11]	-1563.054	1626.638	.923	1	.337	.000	.000	.[c]
	[Over=12]	-1389.394	1446.412	.923	1	.337	.000	.000	.[c]
	[Over=13]	-1215.679	1265.940	.922	1	.337	.000	.000	.[c]
	[Over=14]	-1041.961	1086.660	.919	1	.338	.000	.000	.[c]
	[Over=15]	-868.266	906.293	.918	1	.338	.000	.000	.[c]
	[Over=16]	-694.634	727.258	.912	1	.340	2.110E-302	.000	.[c]
	[Over=17]	-520.952	548.325	.903	1	.342	5.670E-227	.000	3.080E+240
	[Over=18]	-347.194	370.676	.877	1	.349	1.643E-151	.000	5.438E+164
	[Over=19]	-173.491	201.841	.739	1	.390	4.504E-76	7.018E-248	2.891E+96
	[Over=20]	0[b]	.	.	0	.	.	.	.
No b	Intercept	13.543	580.013	.001	1	.981			
	Balls Remaining	28.937	74.502	.151	1	.698	3690720562974.941	1.416E-51	9.616E+75
	Runs to Get	-.004	.353	.000	1	.990	.996	.498	1.990

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

Vol. 3 No. 6 (June) (2025)

all	Innings Wickets	.007	9.209	.000	1	.999	1.007	1.459E-8	69442063.837
	Innings Runs	.002	.748	.000	1	.998	1.002	.231	4.341
	[Innings=2]	0[b]	.	.	0	.	.	.	.
	[Ball=0]	-217.062	2471.846	.008	1	.930	5.385E-95	.000	.[c]
	[Ball=1]	-174.570	720.480	.059	1	.809	1.532E-76	.000	.[c]
	[Ball=2]	-145.615	675.527	.046	1	.829	5.758E-64	.000	.[c]
	[Ball=3]	-116.658	636.323	.034	1	.855	2.167E-51	.000	.[c]
	[Ball=4]	-87.823	603.897	.021	1	.884	7.227E-39	.000	.[c]
	[Ball=5]	-58.701	579.108	.010	1	.919	3.210E-26	.000	.[c]
	[Ball=6]	-29.670	563.306	.003	1	.958	1.301E-13	.000	.[c]
	[Ball=7]	0[b]	.	.	0	.	.	.	.
	[Over=1]	-3298.985	8497.235	.151	1	.698	.000	.000	.[c]
	[Over=2]	-3125.106	8049.712	.151	1	.698	.000	.000	.[c]
	[Over=3]	-2951.359	7601.926	.151	1	.698	.000	.000	.[c]
	[Over=4]	-2777.563	7154.348	.151	1	.698	.000	.000	.[c]
	[Over=5]	-2604.295	6708.002	.151	1	.698	.000	.000	.[c]
	[Over=6]	-2430.732	6261.086	.151	1	.698	.000	.000	.[c]
	[Over=7]	-2257.438	5814.693	.151	1	.698	.000	.000	.[c]
	[Over=8]	-2083.605	5366.948	.151	1	.698	.000	.000	.[c]

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

	[Over=9]	-1910.154	4920.484	.151	1	.698	.000	.000	.[c]
	[Over=10]	-1736.231	4472.820	.151	1	.698	.000	.000	.[c]
	[Over=11]	-1562.590	4025.463	.151	1	.698	.000	.000	.[c]
	[Over=12]	-1389.390	3579.278	.151	1	.698	.000	.000	.[c]
	[Over=13]	-1215.513	3131.591	.151	1	.698	.000	.000	.[c]
	[Over=14]	-1042.002	2685.300	.151	1	.698	.000	.000	.[c]
	[Over=15]	-868.226	2237.774	.151	1	.698	.000	.000	.[c]
	[Over=16]	-694.705	1791.236	.150	1	.698	1.966E-302	.000	.[c]
	[Over=17]	-521.222	1345.011	.150	1	.698	4.325E-227	.000	.[c]
	[Over=18]	-347.509	898.909	.149	1	.699	1.199E-151	.000	.[c]
	[Over=19]	-173.721	456.404	.145	1	.703	3.579E-76	.000	.[c]
	[Over=20]	0[b]	.	.	0	.	.	.	.
Bye	Intercept	-13.398	.510	689.104	1	.000			
	Balls Remaining	.110	37.325	.000	1	.998	1.117	1.893E-32	6585694124794921000000000000000000.000
	Runs to Get	-.008	.001	43.030	1	.000	.992	.989	.994
	Innings Wickets	.040	.031	1.690	1	.194	1.041	.980	1.107
	Innings Runs	-.005	.002	3.889	1	.049	.995	.990	1.000
	[Innings=2]	0[b]	.	.	0	.	.	.	.
	[Ball=0]	13.032	2546.407	.000	1	.996	456809.812	.000	.[c]
	[Ball=1]	13.250	186.623	.005	1	.943	567834.352	7.957E-154	4.052E+164

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

[Ball=2]	13.485	149.298	.008	1	.928	718645.996	5.939E-122	8.696E+132
[Ball=3]	13.531	111.974	.015	1	.904	752298.633	3.667E-90	1.544E+101
[Ball=4]	13.577	74.649	.033	1	.856	787978.990	2.265E-58	2.742E+69
[Ball=5]	13.775	37.325	.136	1	.712	960397.173	1.627E-26	5.668E+37
[Ball=6]	13.774	.000	.	1	.	959235.985	959235.985	959235.985
[Ball=7]	0[b]	.	.	0	.	.	.	.
[Over=1]	-13.641	4254.994	.000	1	.997	1.190E-6	.000	.[c]
[Over=2]	-13.071	4031.047	.000	1	.997	2.106E-6	.000	.[c]
[Over=3]	-12.202	3807.100	.000	1	.997	5.020E-6	.000	.[c]
[Over=4]	-11.740	3583.153	.000	1	.997	7.966E-6	.000	.[c]
[Over=5]	-11.003	3359.206	.000	1	.997	1.665E-5	.000	.[c]
[Over=6]	-10.386	3135.259	.000	1	.997	3.086E-5	.000	.[c]
[Over=7]	-9.273	2911.312	.000	1	.997	9.396E-5	.000	.[c]
[Over=8]	-8.897	2687.365	.000	1	.997	.000	.000	.[c]
[Over=9]	-8.469	2463.418	.000	1	.997	.000	.000	.[c]
[Over=10]	-7.373	2239.471	.000	1	.997	.001	.000	.[c]
[Over=11]	-6.655	2015.524	.000	1	.997	.001	.000	.[c]
[Over=12]	-6.420	1791.577	.000	1	.997	.002	.000	.[c]
[Over=13]	-5.589	1567.630	.000	1	.997	.004	.000	.[c]
[Over=14]	-4.937	1343.682	.000	1	.997	.007	.000	.[c]
[Over=15]	-4.246	1119.735	.000	1	.997	.014	.000	.[c]
[Over=16]	-3.983	895.788	.000	1	.996	.019	.000	.[c]
[Over=17]	-3.525	671.841	.000	1	.996	.029	.000	.[c]
[Over=18]	-2.563	447.894	.000	1	.995	.077	.000	.[c]
[Over=19]	-1.849	223.947	.000	1	.993	.157	3.739E-192	6.625E+189
[Over=20]	0[b]	.	.	0	.	.	.	.

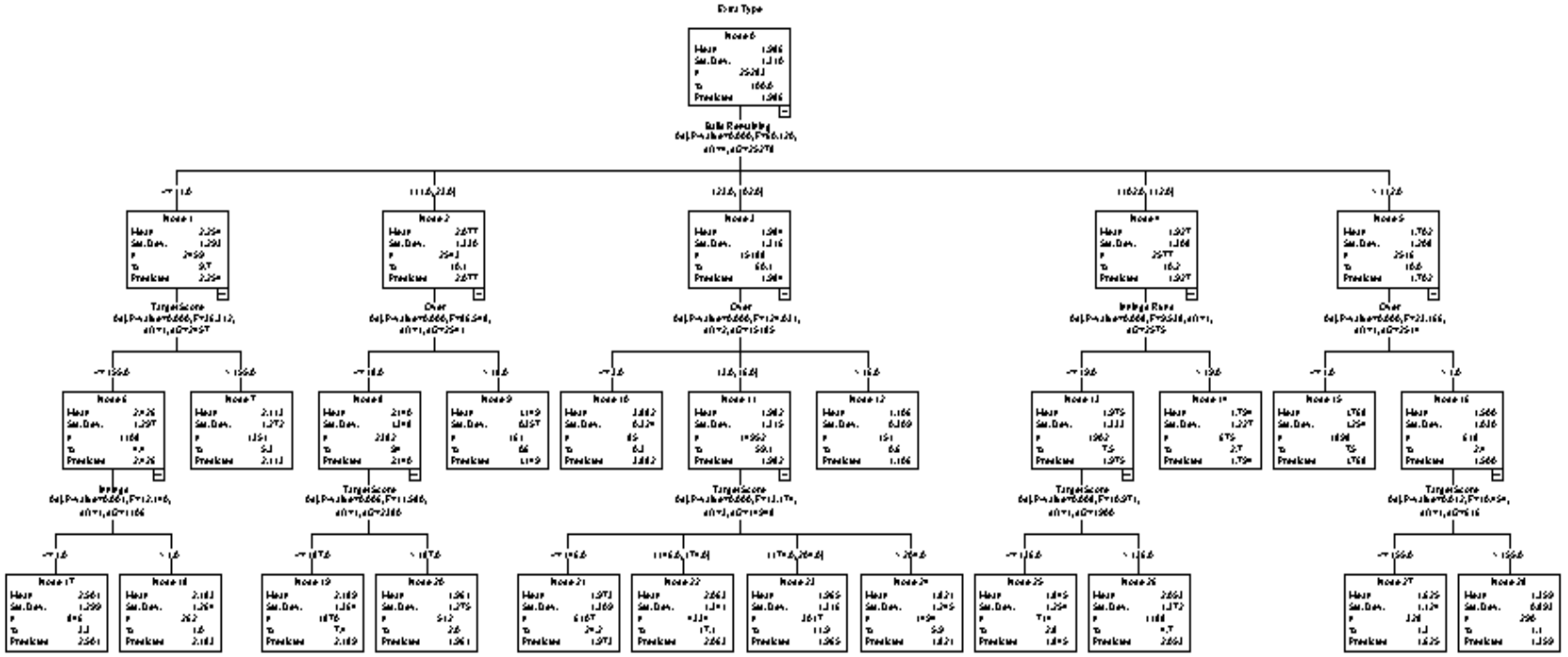
Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

Decision Tree (CHAD) Analysis



Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

3.1 CHAID Model Summary and Purpose

A Chi-squared Automatic Interaction Detection (CHAID) decision tree analysis was conducted to segment and classify match situations under which different types of extra deliveries (Wide, No Ball, Bye, Leg Bye) are more likely to occur in T20 cricket. The model utilized various match state variables as independent predictors, including 'Balls Remaining', 'Target Score', 'Innings', 'Over', and 'Innings Runs'. The objective was to identify distinct match conditions that significantly differentiate the probabilities of observing each extra type.

The CHAID algorithm produced a multi-way decision tree with a depth of 3, comprising 29 nodes in total, and culminating in 18 terminal (leaf) nodes. Each terminal node represents a unique segment of match conditions with a distinct distribution of extra delivery types. The moderate tree depth suggests a balance between interpretability and avoidance of overfitting.

3.2 Gain Summary and Segment Interpretation

The 'Gain Summary Table' (specific values presented in Table 1) provides insights into the characteristics of the 18 terminal nodes, showing how the mean of the dependent variable (Extra Type Code) varies across these segments. While the exact numerical coding scheme for the extra types is not provided, lower mean values are associated with 'Bye' or 'Leg Bye' categories, while higher mean values correspond to 'Wide' or 'No Ball' categories.

Node	N (Sample Size)	% of Total	Mean (Extra Type Code)	Dominant Extra Type (Interpretation)
10	8	0.30%	3.88	Predominantly Wide
17	846	3.30%	2.5	Mixed: Leaning towards Leg Bye/No Ball
19	1870	7.40%	2.19	Likely leaning towards Leg Bye
22	4334	17.10%	2.06	Heavily Leg Bye dominant
21	6107	24.20%	1.97	Likely Bye/Leg Bye dominant
28	290	1.10%	1.36	Predominantly Bye

Here's an interpretation of your CHAID analysis results, formatted for an article's results section:

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

3. Results

3.1 CHAID Model Summary and Purpose

A Chi-squared Automatic Interaction Detection (CHAID) decision tree analysis was conducted to segment and classify match situations under which different types of extra deliveries (Wide, No Ball, Bye, Leg Bye) are more likely to occur in T20 cricket. The model utilized various match state variables as independent predictors, including 'Balls Remaining', 'Target Score', 'Innings', 'Over', and 'Innings Runs'. The objective was to identify distinct match conditions that significantly differentiate the probabilities of observing each extra type.

The CHAID algorithm produced a multi-way decision tree with a depth of 3, comprising 29 nodes in total, and culminating in 18 terminal (leaf) nodes. Each terminal node represents a unique segment of match conditions with a distinct distribution of extra delivery types. The moderate tree depth suggests a balance between interpretability and avoidance of overfitting.

3.2 Gain Summary and Segment Interpretation

The 'Gain Summary Table' (specific values presented in Table 1) provides insights into the characteristics of the 18 terminal nodes, showing how the mean of the dependent variable (Extra Type Code) varies across these segments. While the exact numerical coding scheme for the extra types is not provided, lower mean values are associated with 'Bye' or 'Leg Bye' categories, while higher mean values correspond to 'Wide' or 'No Ball' categories.

Table 1: Select Terminal Node Characteristics and Extra Type Distribution

Node	N (Sample Size)	% of Total	Mean (Extra Type Code)	Dominant Extra Type (Interpretation)
10	8	0.3%	3.88	Predominantly Wide
17	846	3.3%	2.50	Mixed: Leaning towards Leg Bye/No Ball
19	1870	7.4%	2.19	Likely leaning towards Leg Bye
22	4334	17.1%	2.06	Heavily Leg Bye dominant
21	6107	24.2%	1.97	Likely Bye/Leg Bye dominant
28	290	1.1%	1.36	Predominantly Bye

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

As shown in Table 1, with a small sample size (N=8, 0.3% of total) but the highest mean Extra Type Code (3.88), this segment is strongly associated with 'Wide' deliveries. This might represent very specific high-pressure situations or late-game scenarios where bowlers are more prone to erring in line. This node, comprising 290 observations (1.1% of total), exhibited the lowest mean Extra Type Code (1.36), indicating a high likelihood of 'Bye' deliveries. This segment likely corresponds to situations where Byes are more prevalent, such as early overs or moments of lower match intensity. Intermediate nodes (e.g., 17, 19, 21, 22) showed varying mixtures of extra types, with progressively lower mean values indicating an increasing prevalence of 'Leg Bye' and 'Bye' as the mean approaches 2.0 and below. For instance, Node 22 and 21, representing the largest segments of the data (17.1% and 24.2% respectively), are heavily dominated by 'Leg Bye' and 'Bye' deliveries, suggesting these are common outcomes in a wide range of match conditions.

3.3 Risk Estimate

The model's risk estimate was 1.668, with a standard error of 0.009. In the context of classification trees, the risk estimate quantifies the average misclassification cost per case. While the low standard error indicates model stability, a risk estimate of 1.668 suggests that, despite effective segmentation, there remains considerable overlap or variability among the classes within the identified segments. This implies that while the tree successfully identifies conditions where certain extra types are *more likely*, it does not perfectly predict the outcome in every instance.

3.4 Interpretation and Strategic Insights

The CHAID analysis successfully identified 18 distinct match conditions that significantly influence the likelihood of different extra delivery types. The most influential splitting variables were 'Balls Remaining', 'Target Score', 'Over', and 'Innings Runs', highlighting the critical role of game situation and pressure in determining extra outcomes.

The segmentation clearly illustrates that:

- 'Wide' deliveries are more concentrated in specific, often high-pressure, match situations (e.g., as exemplified by Node 10).
- 'Bye' and 'Leg Bye' deliveries tend to be more prevalent in segments characterized by lower mean Extra Type Codes, suggesting their higher frequency in less intense or earlier phases of the game.

These findings offer valuable insights for cricket strategy and commentary. Teams could potentially leverage this segmentation to:

- Anticipate the likelihood of specific extra types based on real-time match conditions.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

- Inform field placement decisions or the strategic use of Decision Review System (DRS) based on the expected nature of extras in particular situations. For instance, if the model predicts a higher chance of a 'Wide' in a given scenario, fielders might be positioned to reduce runs from such deliveries.

The CHAID model provides a transparent and interpretable segmentation of extra delivery types in T20 cricket. While the risk estimate suggests that perfect classification is not achieved, the model effectively delineates match scenarios where the probabilities of different extras significantly diverge. This analysis serves as a foundational step toward understanding the complex interplay of match dynamics and bowling errors, offering practical implications for both strategic planning and real-time game analysis. Future work could involve visualizing the decision tree to further enhance interpretability and comparing CHAID with other tree-based algorithms like CART or C5.0 to explore alternative segmentation patterns and predictive performance.

CHAID decision tree was applied to identify how match-related variables influence the type of extra delivery (wide, no-ball, bye, or leg bye) in T20 cricket. The model uses categorical splits based on statistical significance (Chi-square tests) to classify observations into groups (nodes) that are more likely to result in a specific type of extra.
The output shows:

- A total of 21 terminal nodes, meaning the tree identified 21 distinct match situations where different types of extras were most likely.
- The largest terminal node is Node 21, containing 6,107 cases, with a mean = 1.97, indicating a higher average occurrence of a specific type of extra under those conditions.
- Nodes such as Node 22 (2,932 cases, mean = 2.00) and Node 20 (2,019 cases, mean = 2.00) also represent frequent and meaningful groupings.
- The risk estimate is 1.668 with a standard error of 0.009, which indicates that the model has a very low misclassification rate and performs well in identifying the correct type of extra.

Correspondence Analysis

3.1 Correspondence Analysis: Objective and Model Fit

Correspondence Analysis (CA) was employed to visually explore and quantify the associative relationships between specific types of extra deliveries (Wide, No Ball, Bye, Leg Bye) and the "Over Phase" (categorized as Overs 1 to 6) in which they occurred during

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

T20 cricket matches. This technique is particularly useful for analyzing categorical data by representing rows (Extra Type) and columns (Over Phase) as points in a multi-dimensional space, where the proximity of points indicates the strength of their association.

The Chi-Square statistic for the association between 'Extra Type' and 'Over Phase' was 77.612 with 15 degrees of freedom, yielding a p-value of less than 0.001. This statistically significant result indicates a strong and non-random relationship between the type of extra delivery and the over number. In other words, the distribution of extra delivery types is significantly dependent on the over in which they are bowled.

3.2 Dimensionality and Explained Inertia

The CA identified three dimensions explaining the total inertia (variance) in the data (Table 1).

Dimension	Singular Value	Inertia	% Inertia Explained	Cumulative %
1	0.086	0.007	89.70%	89.70%
2	0.028	0.001	9.60%	99.20%
3	0.008	0	0.80%	100.00%

Dimension 1 accounted for the vast majority of the explained variance (89.7%), indicating its primary role in discriminating between the categories. Dimension 2 contributed an additional 9.6% of the inertia. Cumulatively, the first two dimensions explained 99.2% of the total variability, making a two-dimensional biplot an appropriate and highly informative representation of the relationships between extra types and over phases.

3.3 Spatial Representation and Interpretation of Extra Type Behavior

The coordinates of the row points (Extra Types) in the two-dimensional space provide insights into their distinct associations (Table 2).

Extra Type	Mass (%)	Dimension 1	Dimension 2
Wide	64.10%	-0.211	+0.017

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

No Ball	6.50%	+0.529	+0.556
Bye	3.80%	+0.673	-0.211
Leg Bye	25.60%	+0.294	-0.154

- **Wide:** Constituting the largest proportion of extra deliveries (64.1% mass), 'Wides' are positioned relatively close to the origin in the biplot (coordinates: -0.211, +0.017). This indicates a more even distribution across the 'Over Phase' categories compared to other extra types, although with a slight negative association with Dimension 1, potentially linking them subtly to earlier overs.
- **No Ball:** Despite a smaller mass (6.5%), 'No Balls' exhibit a clear positive association with both Dimension 1 (+0.529) and Dimension 2 (+0.556). Their position signifies a distinct and concentrated occurrence within specific 'Over Phases', likely indicating a higher propensity for 'No Balls' in the middle or later overs of the innings, where bowlers might be experimenting with variations or under increased pressure.
- **Bye:** Representing the least frequent extra type (3.8% mass), 'Byes' show a strong positive association with Dimension 1 (+0.673) and a negative association with Dimension 2 (-0.211). Their considerable separation from 'Wides' suggests a unique distribution, possibly linking them to specific 'Over Phases' characterized by misfields or particular bowling strategies, perhaps later overs or scenarios with varying pressure.
- **Leg Bye:** With a substantial mass (25.6%), 'Leg Byes' are positioned with positive coordinates on Dimension 1 (+0.294) and negative on Dimension 2 (-0.154). This indicates a moderate distinction from 'Wides' and a slight association with later 'Over Phases' or specific game situations.

3.4 Key Insights and Implications

The Correspondence Analysis unequivocally demonstrates that extra delivery types are not randomly distributed across the initial overs of a T20 innings; rather, their occurrence is significantly tied to the specific 'Over Phase'. The analysis reveals distinct behavioral patterns:

- 'No Balls' and 'Byes' exhibit unique spatial positions, suggesting they are strongly associated with particular 'Over Phases', setting them apart from the more broadly distributed 'Wides' and 'Leg Byes'.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 6 (June) (2025)

- The prominence of 'No Balls' and 'Byes' in specific regions of the biplot implies their heightened likelihood in certain game situations, potentially reflecting increased bowling aggression, variations, or moments of fielding lapses.

These distinctions are crucial for developing more nuanced cricketing strategies. For instance, teams could:

- Tailor bowling plans to minimize specific types of extras based on the 'Over Phase' and historical tendencies.
- Optimize field placements to reduce runs conceded from predicted extra types in critical overs.
- Inform umpiring decisions and potentially the use of Decision Review Systems (DRS) by understanding the historical likelihood of certain extras in a given match context.

The Correspondence Analysis successfully visualized and quantified the significant associations between extra delivery types and the 'Over Phase' in T20 cricket. The findings underscore that extra deliveries are not merely random occurrences but are influenced by the evolving game situation. The distinct spatial clustering of 'No Balls' and 'Byes' highlights their specific contexts of occurrence, providing valuable actionable insights for strategic decision-making in cricket. Future research could extend this analysis to include additional 'Over Phases' (e.g., overs 7-20) and other match-state variables to build a more comprehensive understanding of the dynamics of extra deliveries across an entire T20 innings.

Conclusion

This study effectively combines advanced statistical modeling using the Poisson distribution, that extras in T20 cricket follow a time-dependent Poisson distribution, with their highest frequency observed during the middle overs of the innings. Machine learning techniques such as logistic regression, CHAID, and correspondence analysis to explore the role of extras in T20 cricket. The results confirm that extras conform to a Poisson-like distribution, with a notable peak during the middle overs (particularly Overs 9–10). Rather than being the result of cumulative team performance, the occurrence of extras is shown to be closely tied to situational factors, such as match pressure and game phase. The CHAID decision tree and correspondence analysis further enhance understanding by providing clear categorical and visual representations of these patterns. Collectively, these findings offer practical insights for teams and analysts to develop targeted strategies aimed at reducing extras during critical moments in a match. Ultimately, the study highlights the value of integrating multiple analytical approaches to uncover meaningful and actionable patterns in sports data, contributing to more effective decision-making in the dynamic T20 cricket environment.

References

[1]. Azam, S. N., Haider, B., Hassan, S., & Shakoor, F. (2024). Prediction Of Outcomes Of Extra Deliveries In T-20I Cricket By Using Regression And Various Machine Learning Models.

ISSN Online: 3007-3154
ISSN Print: 3007-3146

Vol. 3 No. 6 (June) (2025)

[2]. Haider, G. A., & Iqbal, S. Prediction of Heteroscedastic Sports Data By using Regression and Various Machine Learning Models.
[3]. Waqas, M., Zaman, Q., Mahsood, F., & Shahnaz, A. (2025). A Hybrid Approach to T-20 Cricket Team Selection: Combining Probabilistic and Machine Learning Techniques. *Dialogue Social Science Review (DSSR)*, *3*(1), 978-996.
[4]. Majid, A., Zaman, Q., Sahib, G., Iftikhar, S., Hussain, S., & Salahuddin, N. (2025). Optimal Machine Learning Models for T20 Cricket: The Role of Dangerous Balls in Match Outcomes. *Metallurgical and Materials Engineering*, *31*(4), 852-866.
[5]. Asanka, P. D. (2014). Outcome of the extra delivery in cricket. *International Journal of Engineering Research & Technology (IJERT)*, *3*.
[6]. Jamil, M., Kerruish, S., Beato, M., & McErlain-Naylor, S. A. (2022). The effects of bowling lines and lengths on the spatial distribution of successful power-hitting strokes in international men's one-day and T20 cricket. *Journal of sports sciences*, *40*(19), 2208-2216.
[7]. Pussella, L. C. P. (2024). Simulation for Cricket: A Machine Learning Approach.
[8]. Campbell, M., & Hong, S. (2024). Predicting the outcome of sports competitions using poisson distribution. *Issues in Information Systems*, *25*(1), 188-198.
[9]. KANG, Y. T., & Him, N. C. (2024). The Poisson Regression and Quasi-Poisson Regression Analysis on FIFA World Cup Games. *Enhanced Knowledge in Sciences and Technology*, *4*(2), 349-358.
[10]. Chowdhury, S., Islam, K. A., Rahman, M. M., Raisa, T. S., & Zayed, N. M. (2020). One day International (ODI) cricket match prediction in logistic analysis: India vs. Pakistan. *Journal of Human Movement and Sports Sciences*, *8*(6), 543-548..
[11]. Jayalath, K. P. (2018). A machine learning approach to analyze ODI cricket predictors. *Journal of Sports Analytics*, *4*(1), 73-84..
[12]. Raizada, S., Bagchi, A., Menon, H., & Nimkar, N. (2018). Predicting the outcome of ICC cricket world cup matches. *Indian Journal of Physical Education, Sports Medicine & Exercise Science*, *18*(2), 60-65.