Impact of Extras and Top Individual Scores on Team Totals in Asia Cup Cricket

Sai Pranav Sripathi

2025-09-20

This paper investigates whether extra runs—awarded due to bowling errors—meaningfully impact a team's total score in Asia Cup cricket matches (1984–2022). Applying simple linear regression, the analysis finds that extras show only a weak correlation with total runs scored. In search of a stronger predictor, we also examine the effect of a team's highest individual batting score using the same approach. Results demonstrate that while extras offer a modest contribution, the highest individual score is strongly associated with total team runs. These findings highlight that star player performances are much more influential than extras in boosting team outcomes, as detailed in the sections that follow.

1 Introduction

Cricket matches are often decided not just by team effort, but by unique factors such as the number of extra runs and standout individual performances. In high-stakes tournaments like the Asia Cup, understanding what actually drives team scoring is crucial for coaches, players, and analysts alike. The central question of this paper is: *How do extra runs and a team's highest individual batting score each affect the total runs scored in a match?*

This question matters because team strategies frequently target the reduction of extras, and commentators regularly highlight the impact of exceptional single-player innings. Prior research has examined statistical relationships in cricket, but much of it focuses on general predictors of match outcomes or player averages rather than quantifying the effects of extras and individual highs on team totals.

In this analysis, we use match-level data from the Asia Cup (1984–2022) and apply simple linear regression to both extras and highest individual scores as predictors of total runs seperately. We find that extras are only weakly related to team scoring, whereas a team's top individual score strongly influences their total runs.

The remainder of this paper describes the dataset and variables used, details the methods of analysis, presents and interprets the main results, and discusses the strategic and statistical implications of these findings.

The remainder of this paper is as follows. Section 2 focuses on describing the Asia Cup dataset and its principal variables. Section 3 outlines the methods used for analyzing the effects of extras and highest individual scores on team totals. Section 4 presents the main results and interprets their meaning for understanding cricket scoring. Finally, Section 5 discusses the broader strategic significance of these findings.

2 Data and Variables

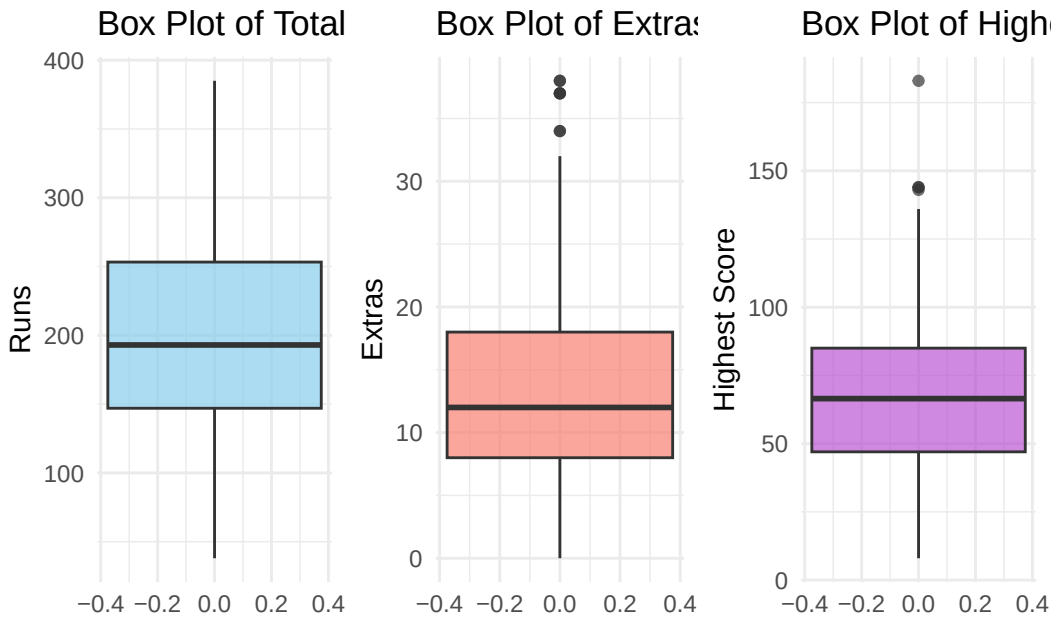
This analysis uses match-level data from the Asia Cup cricket tournaments, accessible at (SCORENetwork 2022) Data Repository. The dataset contains a record for every match played from the inaugural Asia Cup in 1984 up to 2022, now held every alternate year in various Asian host cities.

The principal variables used to answer the research question are:

Run Scored: Total runs scored by each team in a match (response variable).

Extras: Extra runs scored due to bowling errors (wides, no-balls, byes, leg byes).

Highest Score: Highest individual batting score for each team in a match (key predictor).



Additional variables include team name, opponent name, host venue, year, toss outcome, selection after toss, total fours and sixes, match result, and extras given up.

Sample size is approximately 250 team-match records. Each row represents one team's performance in one Asia Cup match. To ensure data quality and reliable results, I checked all main variables for missing (null or empty) values and excluded any rows with incomplete data from the analysis. This step prevents biased results and ensures that all summaries and regression models are based on complete and valid match records.As part of my data preparation, I checked each column for null or empty values and excluded any rows with missing data from the analysis to ensure accuracy and avoid bias in the results.

While the Asia Cup features some of the most cricket-intensive nations globally—India, Pakistan, Sri Lanka, Bangladesh, Afghanistan—other prominent cricket countries such as Australia, England, and West Indies are not included, as they do not participate in the Asia Cup. This restricts generalizability: the analysis applies best to Asian teams, which comprise the majority of international cricket matches, but may not extend to all cricket-playing nations.

To illuminate possible relationships, summary statistics and scatter plots are presented in the results section; for example, histograms of Run Scored and Highest Individual Score, and scatter plots showing their association.

In summary, this dataset is well suited to addressing the question of how extras and highest scores affect team total runs in high-level cricket, though the findings are not fully generalizable to all international cricket contexts.

3 Methods

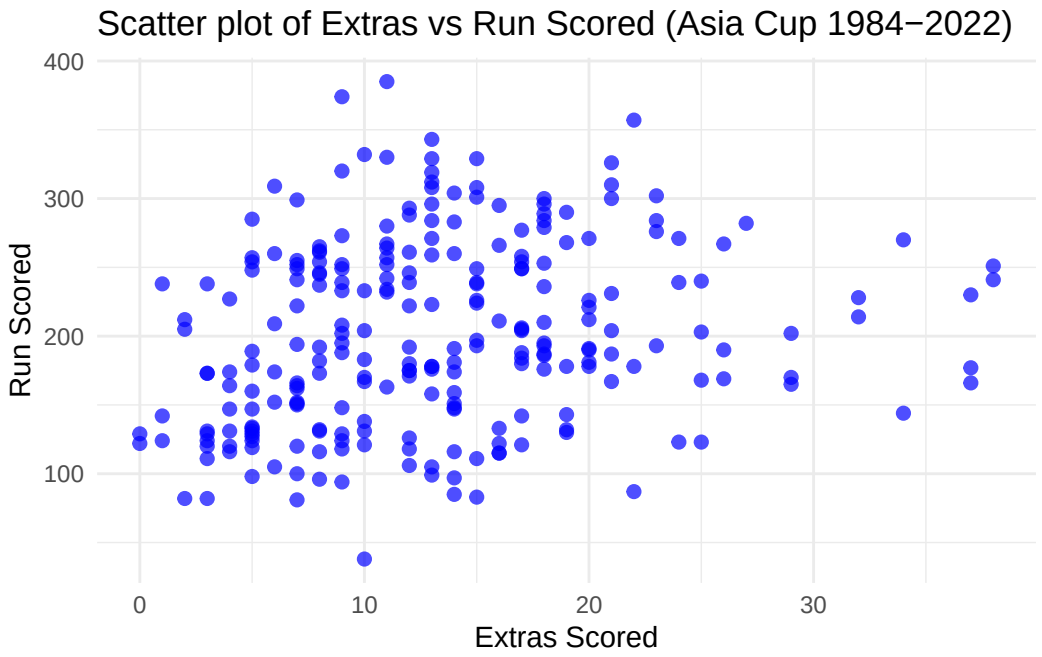
To examine how extras scored and highest individual batting score each influence team totals, I fit two separate simple linear regression models to the Asia Cup data.

First Regression: Effect of Extras Scored

Let $RunScored_i$ denote total runs scored by team i, and $ExtrasScored_i$ the number of extras recorded in match i. The model is:

$$RunScored_i = \beta_0 + \beta_1 * ExtrasScored_i + \epsilon_i$$

where β_0 is the intercept, β_1 measures the expected change in total runs per additional extra run, and ϵ_i is the error term for match i.



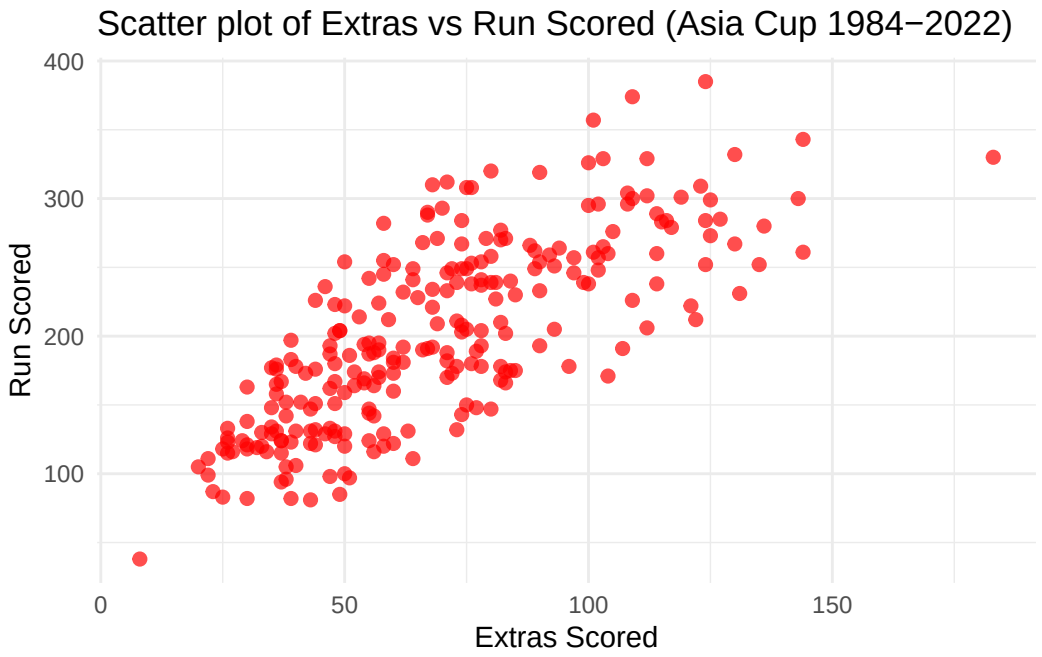
A scatter plot of extras scored versus total runs scored across all Asia Cup matches from 1984 to 2022 reveals a modest but statistically significant positive relationship. In this visualization, each point represents a team's totals for a given match, with a fitted regression line highlighting the association. The linear regression output indicates that, on average, each additional run given as an extra is associated with an increase of approximately 1.88 runs in the team's total score (coefficient = 1.88, $p < 0.001$). However, the adjusted R2R2 value is about 0.04, meaning that extras account for only a small fraction (~4%) of the variation in total scores. While reducing extras can marginally benefit overall performance, these findings suggest that many other factors play larger roles in determining a team's total runs.

Second Regression: Effect of Highest Individual Score

Let $HighestScore_i$ be the highest individual run total for team i in match. The model is:

$$RunScored_i = \beta_0 + \beta_1 * HighestScore_i + \epsilon_i$$

Here, β_0 is the intercept, β_1 captures the effect of the top individual score, and ϵ_i is the unexplained variation for match.



Parameter estimation is conducted using the `lm()` function in the R programming language (R Core Team 2024) with data visualization and manipulation using ggplot2 and dplyr (H. Wickham 2016; Hadley Wickham et al. 2023). Models assume independent match outcomes, additive linear relationships, and normally distributed errors with constant variance.

A main limitation of this approach is that only one predictor is included per model; additional match factors that may affect total runs (e.g., team strength, opposition quality) are not considered here and could introduce omitted variable bias. These limitations are discussed further in the Discussion section.

4 Results

We report the parameter estimates and model summaries for both regression analyses:
(1) the impact of extras scored on total runs, and
(2) the impact of highest individual score on total runs.

1. **Regression of Total Runs on Extras Scored**

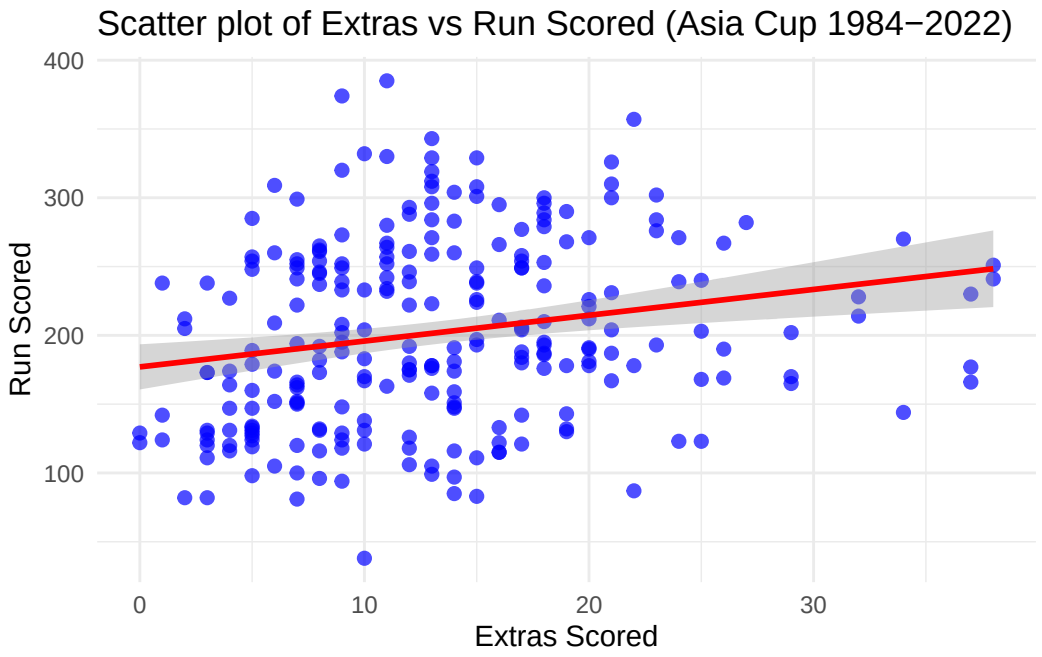
   The model fit and the estimated coefficients are:

Table 1: Estimated coefficients from the regression of total runs on extras scored.

	Parameter	Estimate	Standard.error	t.value	p.value
(Intercept)	β_0 (Intercept)	177.07	8.34	21.24	<0.001
Extras.Scored	β_1 (Slope)	1.88	0.55	3.44	<0.001

Interpretation:

- The slope β_1 represents the expected change in total runs with each additional extra run scored.
- The p-value tests the null hypothesis that this effect is zero. If $p < 0.05$, the relationship is considered statistically significant at the conventional 5% level.
- The R-squared value (found in `summary(fit_extras)$r.squared`) quantifies how much of the variation in total runs is explained by extras alone.



2. Regression of Total Runs on Highest Individual Score

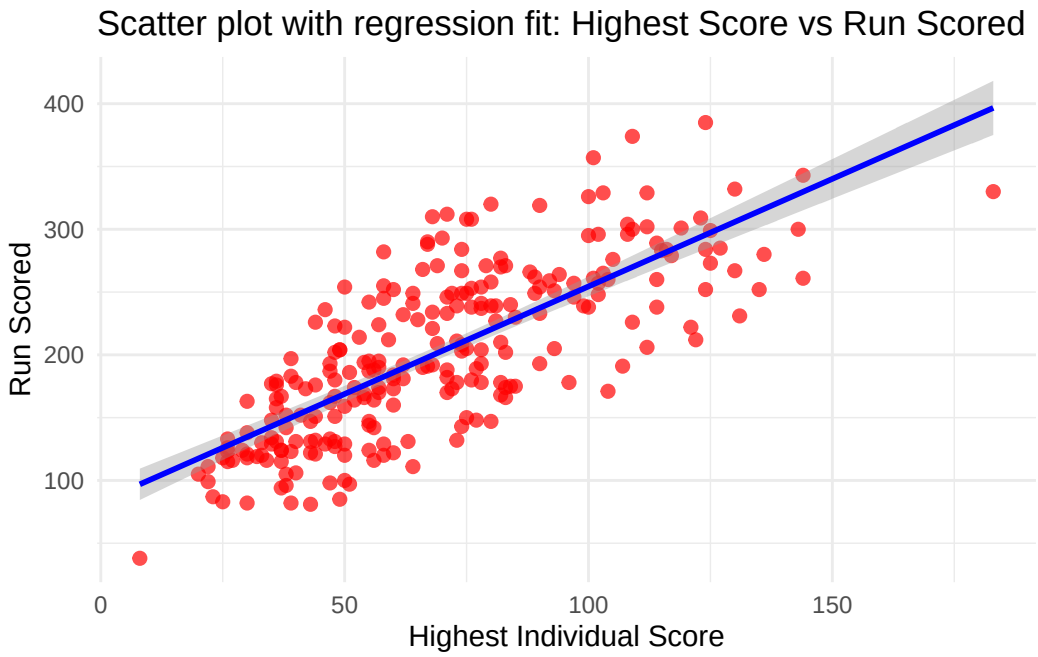
The model fit and the estimated coefficients are:

Table 2: Estimated coefficients from the regression of total runs on highest individual score.

	Parameter	Estimate	Standard.error	t.value	p.value
(Intercept)	β_0 (Intercept)	83.20	7.02	11.85	<0.001
Highest.Score	β_1 (Slope)	1.71	0.09	18.40	<0.001

Interpretation:

- The slope β_1 represents the expected change in total runs for each additional run scored by the team's top batter.
- The p-value and standard error provide evidence about precision and statistical significance.
- R-squared (see `summary(fit_highscore)$r.squared`) indicates how much variation in total runs is explained by highest score alone.



5 Discussion

This study used simple linear regressions to measure how extras scored and a team's highest individual batting score each affect total runs in Asia Cup matches. Results show that while extras provide a statistically significant but modest boost to team totals, the highest individual

score has a stronger and more consistent influence. Thus, top batting performances outweigh the effect of extras in match outcomes.

Key limitations include the restricted dataset (Asia Cup matches only) and the absence of other match variables or tournament contexts. Our linear approach may not capture all complexities of cricket scoring. Future work should incorporate data from non-Asian teams, test additional predictors, and explore non-linear or interaction effects.

In conclusion, cricket teams benefit more from exceptional individual batters than from minimizing extras, guiding both coaching strategies and further research.

References

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

SCORENetwork. 2022. *SCORE Sports Data Repository*. https://data.scorenetwork.org/.

Wickham, H. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.