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 scored due to bowling errors, meaningfully impact a team's total score in cricket matches. 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

The results of a cricket match depends on a lot of factors, such as extras scored, standout individual performances etc. Tournaments with high stakes like Asia Cup, its crucial to understand what drives a team's success. Scoring is a crucial aspect for coaches, players and analysts alike. The central question of this paper is: How do extras scored and a team's highest individual score in a match affect the total score of the team. **Extras are runs awarded to the batting team as a result of rule infringements or errors committed by the bowling team.**

This is a significant issue, as the team strategies would generally involve reducing the extras, and experts often highlight the impact of an extraordinary performance of a single player. Existing research focus on statistical relations alone, but main concentration is on the general predictors of match outcomes or on players averages, rather on the effects of extras and individual performances on team totals.

In this analysis, we analyse match-level data of the Asia Cup from 1984 to 2022, and apply simple linear regression to regress total score with extras scored and highest individual score

separately. We discover that extras are weakly related to the total team score, whereas the highest individual score has a major impact on their total scores.

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

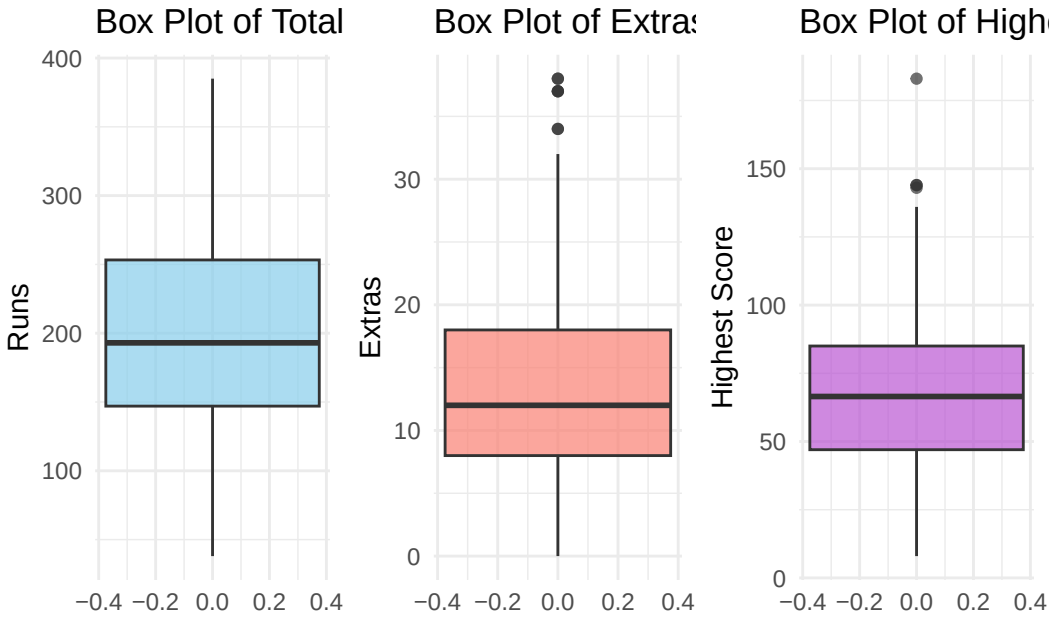
This analysis uses match-level data from the Asia Cup tournaments, accessible at (SCORENetwork 2022) Data Repository. The dataset contains a record for every match played from the inaugural Asia Cup in 1984 up to 2022, now held every alternate year in various Asian host cities.

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.

The sample is approximately 250 matches records. Each row represents a match and the host team's performance in it. To ensure data quality and reliable results, I checked the main variables for missing (null or empty) values and there were none. This step prevents biased results and ensures that all summaries and regression models are based on complete and valid match records.

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 makes generalization difficult: the analysis applies best to Asian teams who's data is being analysed, which comprise the majority of international cricket matches, but may not extend to all cricket-playing nations.

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 generalized to all international cricket contexts.

3 Methods

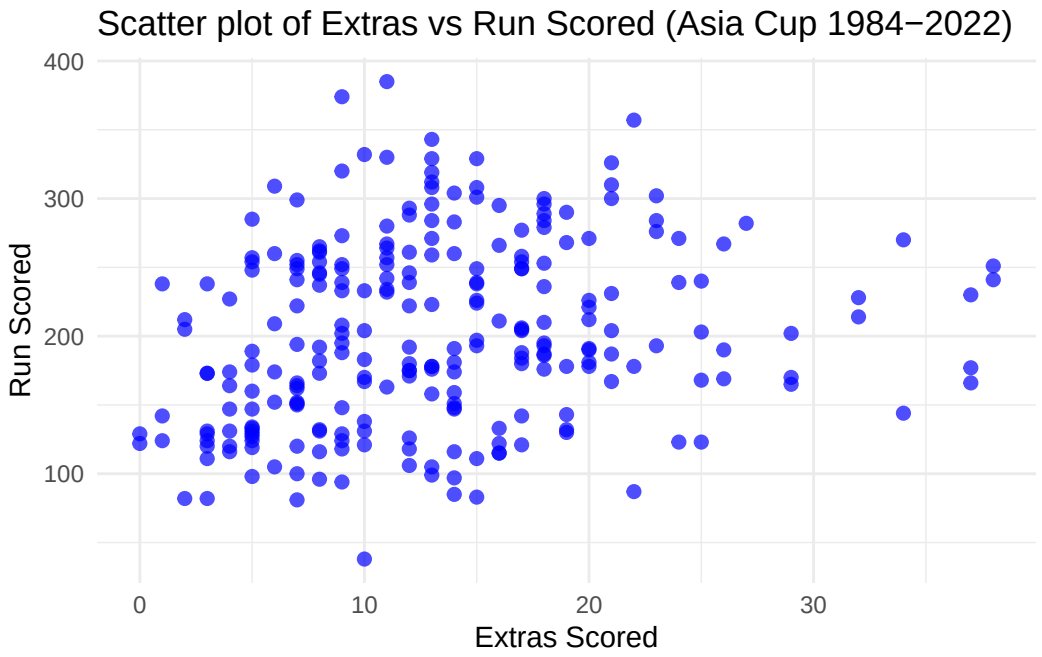
The affect of extras scored and highest individual batting score on teams total is examined by two seperate simple linear regression models to the entire Asia Cup data available.

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.



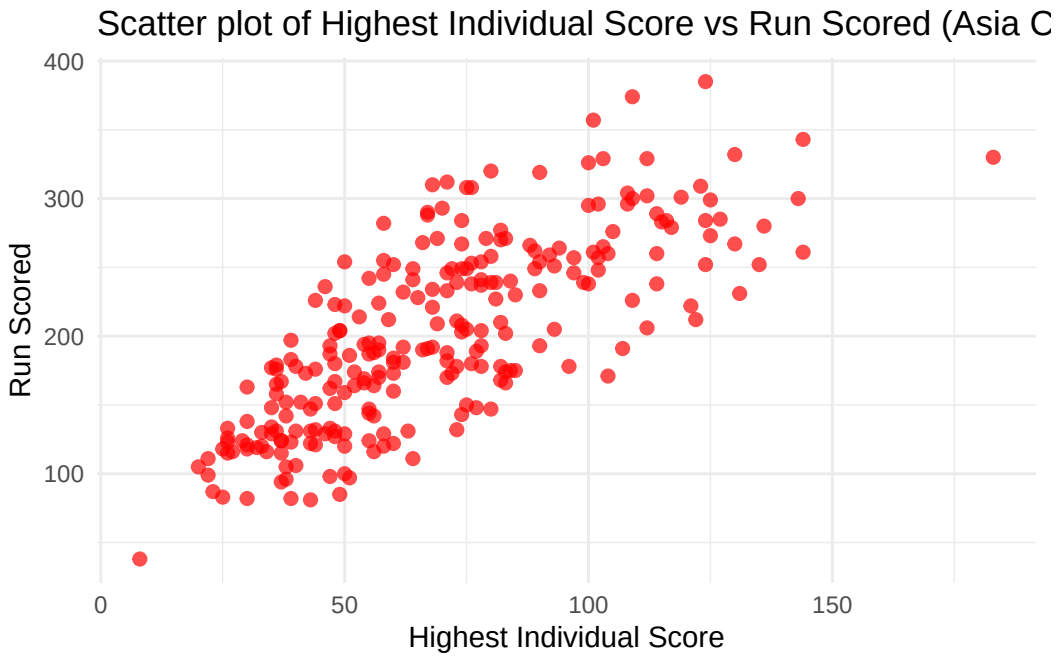
A scatter plot of extras scored versus total runs scored across all Asia Cup matches from 1984 to 2022 reveals a modest/low but statistically significant positive relationship. In the visualization, each point represents a team's totals for a given match with the extras scored on corresponding axes.

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.



A scatter plot of highest individual score versus total runs scored across all Asia Cup matches from 1984 to 2022 reveals a high and significant positive relationship. In the visualization, each point represents a team's totals for a given match with the highest individual score on corresponding axes.

Parameters estimation is done 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 the match outcomes are independent.

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**

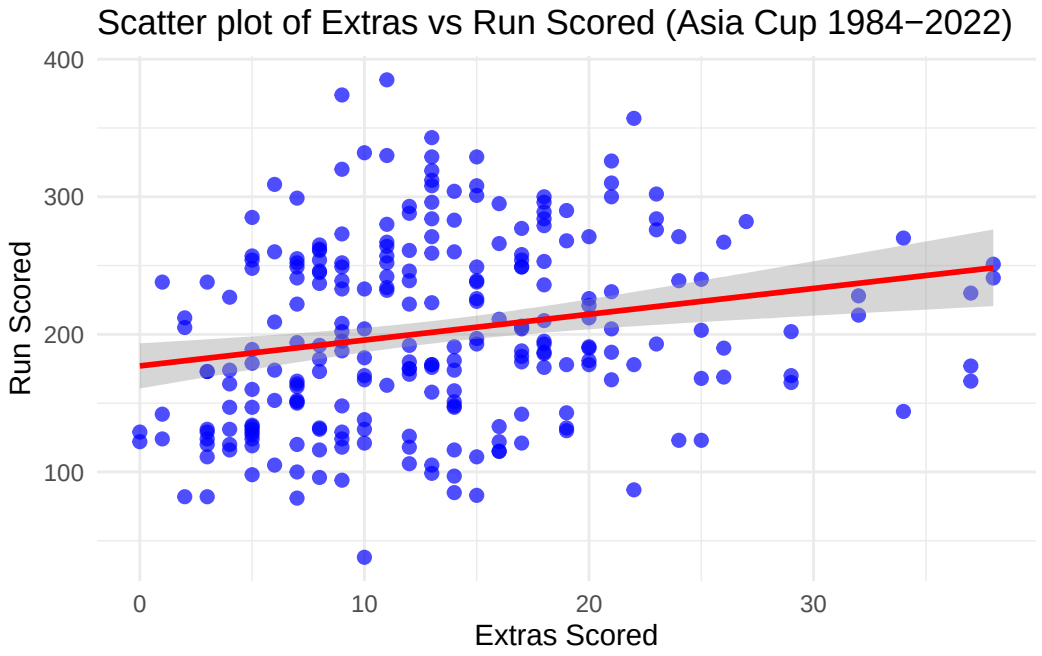
   The model fit and the estimated coefficients are:

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 intercept estimate $\hat{\beta_0} = b_0 = 177.07$ represents that the expected total runs in a match is 177 when the extras scored is 0, it is statistically significant as the standard error is 8.34 which is $< 5\%$
- The slope estimate $\hat{\beta_1} = b_1 = 1.88$ represents the expected change in total runs with each additional extra run scored, i .e, it is expected that an increase of 1.88 run for every additional extra scored in a match.
- 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. Which mean that there is a good amount of evidence that extras scored has a positive effect on the total score in a match.
- This analysis prove that the disciplined bowling and fielding to avoid or minimize extras will boost a team's chance of conceding high scores and hence would have a better shot at winning.



2. Regression of Total Runs on Highest Individual Score

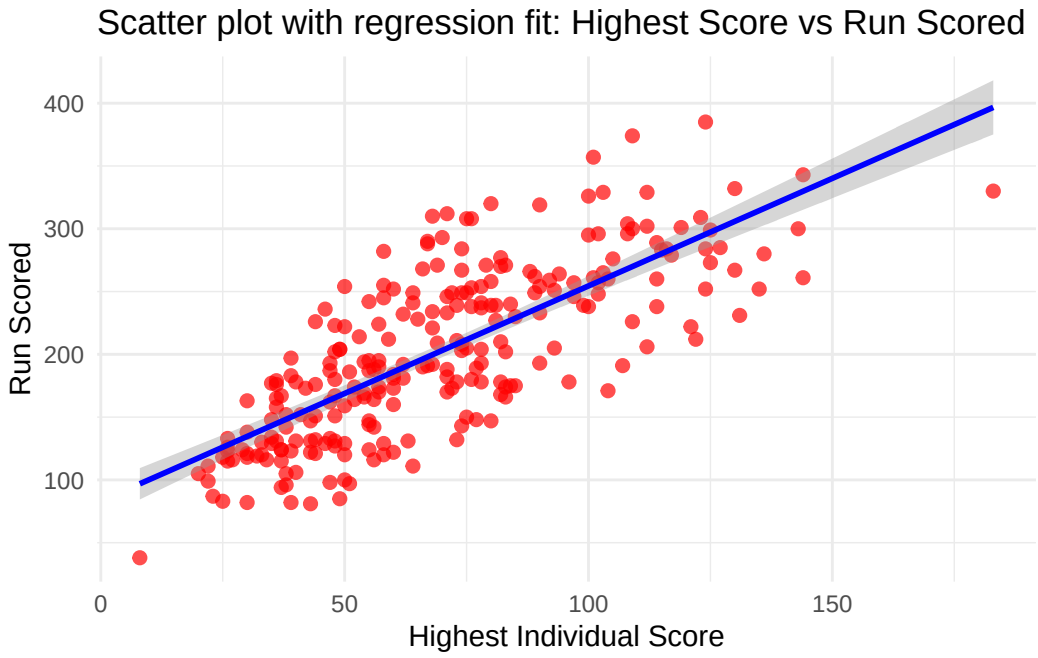
The model fit and the estimated coefficients are:

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 intercept estimate $\hat{\beta}_0 = b_0 = 83.20$ indicate that the expected team total would be 83.2 if the highest individual score is 0, which does not make a lot of sense. But based on the available data and the regression fit, this is interpreted.
- The slope estimate $\hat{\beta}_1 = b_1 = 1.71$ represents the expected change in total runs for each additional run scored by the team's top batter, i.e, with every additional run scored by the top batter in a match, the teams total is expected to rise by 1.71 runs.
- The standard errors of both the estimates are quite low.
- The p-value and standard error provide evidence about precision and statistical significance and not just any random chance but a real and consistent effect in the cricket matches.



5 Discussion

This study used simple linear regressions to measure and quantify 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. In contrast the highest individual score has a stronger and more consistent influence on the total score of a team in the matches. These findings show that the teams with high score achievers have an improved chances to score more and thus a better chance at victory while minimizing the number of extras they concede.

Key limitations include the restricted dataset (Asia Cup matches only) and the absence of other match variables or tournament contexts. Along with the other key aspects like pitch condition, moisture in the air, team strengths have not been considered. 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. With other data points like the pitch, weather, partnerships would tend to give more insights on the data.

In conclusion, the evidences suggest that exceptional individual batters who can produce high scores would yield more benefits to the teams scores than focusing exclusively on reducing extras. These insights have practical impacts while forming strategies and coaching any team for the tournaments.

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.