

# A Predictive Model for ODI World Cup Matches

Krishna Kumar, Srihari Srinivasan

APR 30, 2023

## Introduction

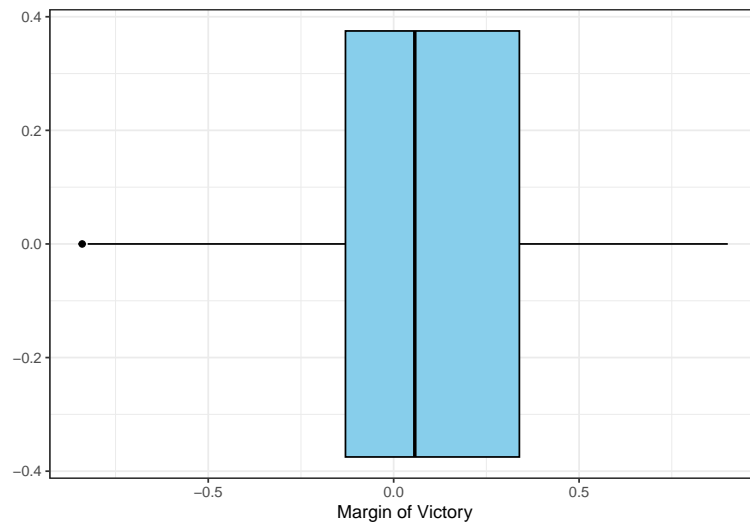
Given that this is a One Day International (ODI) World Cup year, our analysis of batting and bowling performance metrics on the margin of victory, `mov`, is focused on prior World Cup matches. With knowledge of how the ODI format has changed since the introduction of the shorter, Twenty20 (T20) format, particularly the 2007 T20 World Cup, we have limited our data to the last five ODI World Cups (2003, 2007, 2011, 2015, and 2019). After wrangling six batting metrics and four bowling metrics from 4774 individual player performances, we are looking to build a predictive model and gain insight as to which metrics, if any, have an effect on `mov`. In doing so, while this may be beyond the scope of this project, we hope to ultimately use our model to predict the outcome of the 2023 ODI World Cup.

## Exploratory Analysis

### Response Variable

`mov` is the difference between the target set by the team batting first and the total that the chasing team achieved. For example, in a 2003 match between England and Pakistan, England scored 246 runs, setting 247 as the target for Pakistan to chase. They, however, were bowled out for 134, resulting in England winning by 112 runs. Therefore, the `mov` for this match is calculated as  $winner\_margin/target = 112/247 = 0.453$ .

The figure below is a boxplot of `mov` values.



```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.8400 -0.1300   0.0570   0.0709  0.3390   0.9010
```

From the figure and summary, we can see that `mov` is approximately normally distributed on 7.1%. The single outlier is a 2011 match between Kenya and New Zealand in which Kenya lost by 84.0%. While large `mov` values can be attributed to a blowout, the extreme values probably occurred due to a wide skill gap between two teams. New Zealand, for example, is an established cricketing nation with a strong, experienced team compared to Kenya.

## **Predictor Variables**

Batting metrics:

1. `to_runs_pct` - percentage of total runs scored by the top order
2. `mo_runs_pct` - percentage of total runs scored by the middle order
3. `to_mins_pct` - percentage of total time spent in crease by the top order
4. `mo_mins_pct` - percentage of total time spent in crease by the middle order
5. `to_bf_pct` - percentage of total balls faced by the top order
6. `mo_bf_pct` - percentage of total balls faced by the middle order
7. `pct_4s` - percentage of total runs that are 4s
8. `pct_6s` - percentage of total runs that are 6s
9. `to_sr` - average strike rate of the top order
10. `mo_sr` - average strike rate of the middle order

## **Model Development**

## **Model Analysis**

## **Conclusion**