Indian Premier League Cricket

Matt Stuart[1,2], Hassan Raffique[3], Leigha DeRango[1]
Gregory J. Matthews[1,2]

[1] Department of Mathematics and Statistics, Loyola University Chicago, Chicago, IL, U
[2] Center for Data Science and Consulting, Loyola University Chicago, Chicago, IL, U
[3] Univsersiry of Iowa, Iowa City, IA
[+] Corresponding: mstaurt@luc.edu

Abstract

Wicked Googly

Keywords: Cricket

1 Introduction

openWAR and cricWAR.

In the game of cricket, the number of runs scored on a particular pitch typically ranges between 0 and 6 (though theoretically values larger than 6 are possible, they are rare and do not occur at all in our particular data set).

2 Data

season	n
2015	8
2016	8
2017	8
2018	8
2019	8
2021	8
2022	10

We have data from the Indian Premier Leauge (IPL) consisting of 102490 pitches from the 2015 - 2022 seasons. From 2015 - 2021, the league had 8 teams followed by 10 teams in the 2022 season.

season	n
2015	13641
2016	14096
2017	13849
2018	14286
2019	14293
2021	14413
2022	17912

Runs in cricket are either scored by running back and forth between the wickets once the ball is put into play (generally resulting in 1 or 2 runs, but theoretically any value is possible). In addition, a ball that is hit in the air over the boundary (termed a “boundary”) is worth 6 runs and if the ball rolls to the boundary or bounces in the field of play and then clears the boundary this is worth 4 runs (termed a “boundary 4”). As a result the distribution of runs scored on a a particular pitch has large peaks are 0 and 1 with a big drop off from 1 to 2.

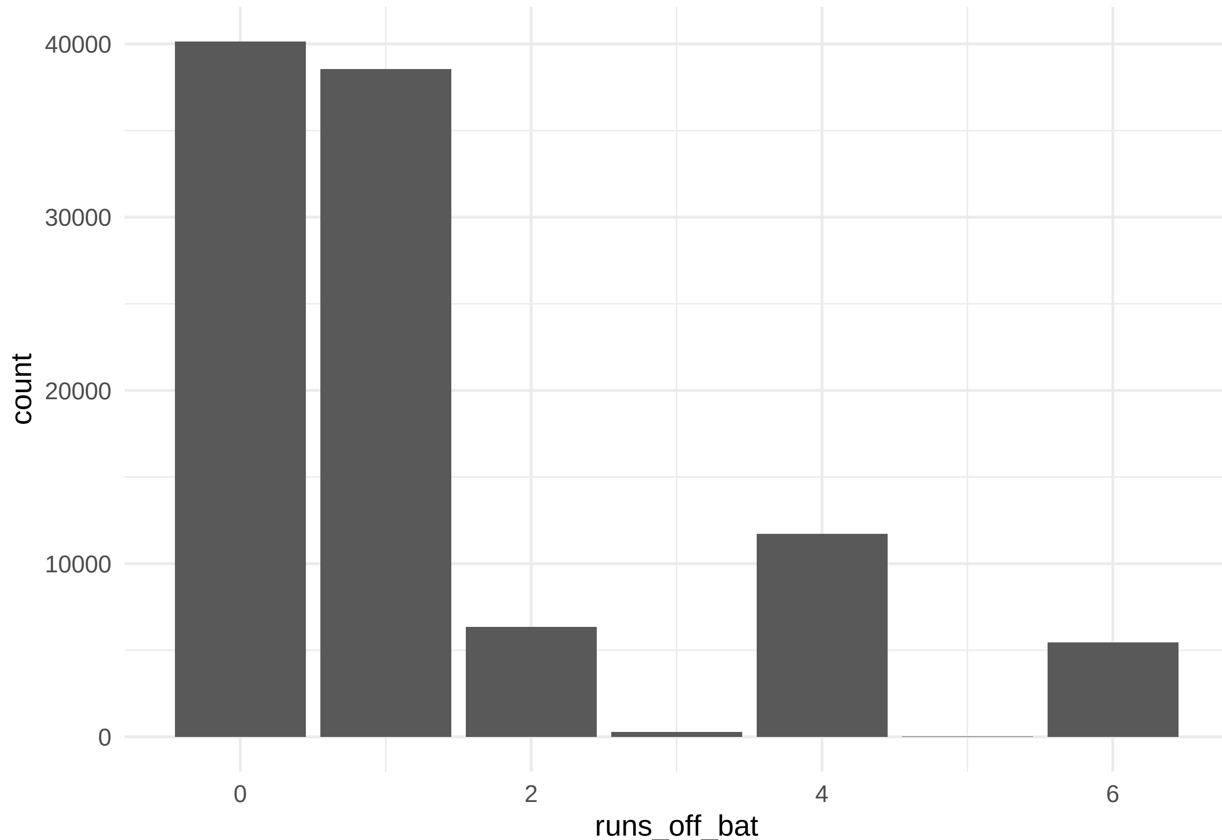

Figure 1: Bar Plot of the number of runs scored of a particular pitch

Values of 3 and 5 are extremely rare accounting for only 0.27% and 0.02% of values across all pitches in our data set. Values of 4 and 6 spike because of boundaries and boundary fours and together account for 16.75% of all values.

Figure 1 shows a bar plot.

3 Models

Let r_i be the number of runs scored on the i-th ball with $i = 1, \cdots, n$ and $r_i \in \{0, 1, \cdots, 6\}$. Then define $\mathbf{y_i} = (y_{i0}, \ldots, y_{i6})$ where $y_{ij} = I(r_i = j), \forall j = 0, \cdots, 6$ and $I(.)$ is the indicator function. We then model:

$$\mathbf{y_i} \sim MN(1, \mathbf{p_i})$$

where $\mathbf{p_i} = (p_{i0}, \ldots, p_{i6})$, p_{ij} is the probability that j runes are scored on the i-th ball, and $\sum_{j=0}^{6} p_{ij} = 1$ for all i.

$$log\left(\frac{p_j}{p_0}\right) = \beta_{0j} + \mathbf{X}\beta_j + b_{j,batter} + b_{j,bowler} + b_{j,runner}$$

for $j = 1, \ldots, 6$ where β_{0j} is the intercept of the j-th linear component, β_j is a P-dimensional vector containing the regression coefficients for the fixed effects, $\mathbf{X}$ is the matrix of covariates, and $b_{j,batter}, b_{j,bowler}, b_{j,runner}$ are random effects for the batter, bowler, and runner, respectively.

3.1 Priors

$\beta_{j0} \sim N(0,1)$ for $j = 1, \ldots, 6$ and $\beta_{jp} \sim N(0,1)$ for $p = 1, \ldots, P$ and $j = 1, \ldots, 6$.

$b_{j,batter} \sim N(0, \sigma^2_{batter})$ for all batters. $b_{j,bowler} \sim N(0, \sigma^2_{bowler})$ for all bowlers. $b_{j,runner} \sim N(0, \sigma^2_{runner})$ for all runners.

$\sigma^2_{batter}, \sigma^2_{bowler}, \sigma^2_{runner} \sim Inv - \chi^2(something)$

4 Results

5 Discsusson, Future work and conclusions

Acknowledgements

Supplementary Material

All code for reproducing the analyses in this paper is publicly available at https://github.com/gjm112/cricketIPL

6 References