arXiv:2512.19740v1 [cs.LG] 17 Dec 2025

2025年亚洲杯：用于板球分析的结构化 T20 比赛级数据集与探索性分析

Kousar Raza[1*] **Faizan Ali**[2]

[1]伊朗伊斯法罕大学应用数学与计算机科学系[2]巴基斯坦贾姆肖罗信德大学数学与计算机科学研究所

摘要

本文介绍了一个关于2025年亚洲杯 T20 板球锦标赛的结构化综合数据集，旨在促进体育分析领域的数据驱动研究。该数据集涵盖锦标赛全部 19 场比赛的记录，包含 61 个变量，包括球队得分、 विकेट数、强打阶段统计、边界球数量、掷币决策、比赛场地和球员个人亮点。为展示其分析价值，我们开展了探索性数据分析，重点考察球队表现指标、边界分布和得分模式。该数据集通过 Zenodo 公开发布，采用知识共享署名 4.0 许可，以支持可复现性，以及板球分析、预测建模和战略决策方面的进一步研究。本研究提供了一个开放的机器可读基准数据集，助力推进板球分析研究。

关键词 s: 板球分析、体育数据集、T20、2025年亚洲杯、开放科学

1 引言

板球分析日益依赖高质量数据集，以支持统计分析和机器学习应用。虽然主要锦标赛通常可以获取逐球数据，但针对特定国际赛事、可公开访问且文档完善的比赛级数据集仍然十分有限。2025年亚洲杯T20锦标赛为分析高压环境下的短赛制板球提供了宝贵背景。本文介绍了**2025年亚洲杯比赛级数据集**，旨在支持体育分析领域的可复现研究。

*ORCID：0009-0007-5298-4469

2 数据集描述

该数据集以单个 CSV 文件 `asia cup 2025 complete dataset.csv` 的形式发布，包含锦标赛全部19 场比赛的记录，每场比赛包含61 个属性。

2.1 数据结构与变量

表 1 总结了数据集中包含的主要变量类别。

类别	代表性变量
比赛识别	比赛编号，Date，比赛场地，系列赛
球队	球队1、球队2、出场十一人
掷硬币信息	掷硬币获胜方，掷硬币决定
比赛结果	结果、获胜优势、全场最佳球员
球队表现	总得分，wकिट数，局数，额外得分
强打阶段统计	强打阶段得分、强打阶段投球数
边界得分数据	四分球、六分球
赛事背景	阶段、组别、比赛形式

表 1：2025年亚洲杯数据集中的变量类别概览

2.2 数据来源与收集

比赛数据收集自**ESPNcricinfo** 上公开提供的记分卡，并与**亚洲板球理事会（ACC）**发布的官方锦标赛信息进行交叉核对。数据处理包括人工核验、实体名称标准化，以及跨来源的一致性检查。

3 探索性分析

本次探索性分析旨在展示该数据集的分析潜力，重点关注掷硬币结果、球队击球表现和边界分布。

3.1 掷硬币结果的影响分析

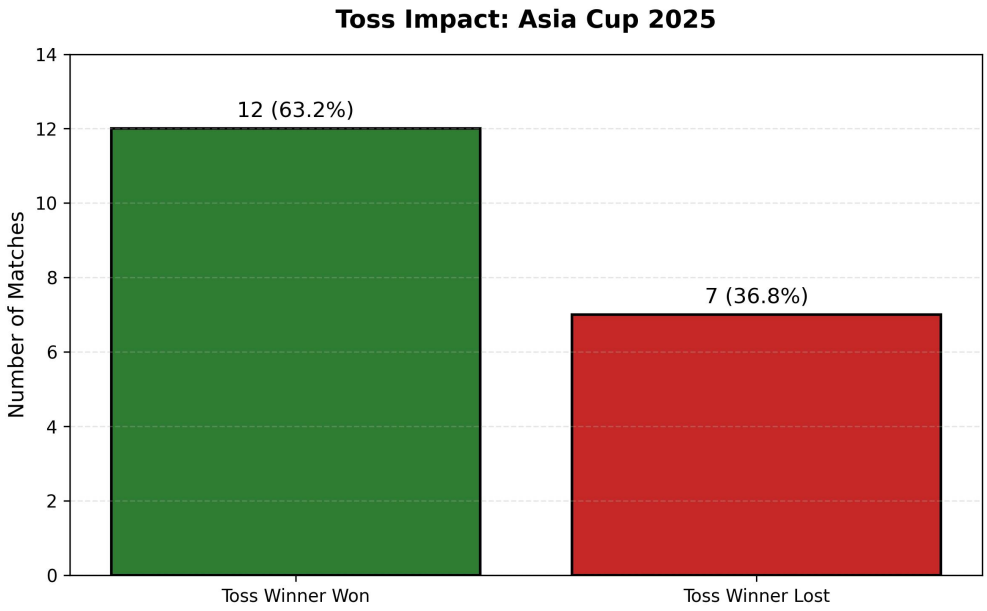

图1：2025年亚洲杯锦标赛中掷硬币结果对比赛结果的影响。

3.2 球队击球表现

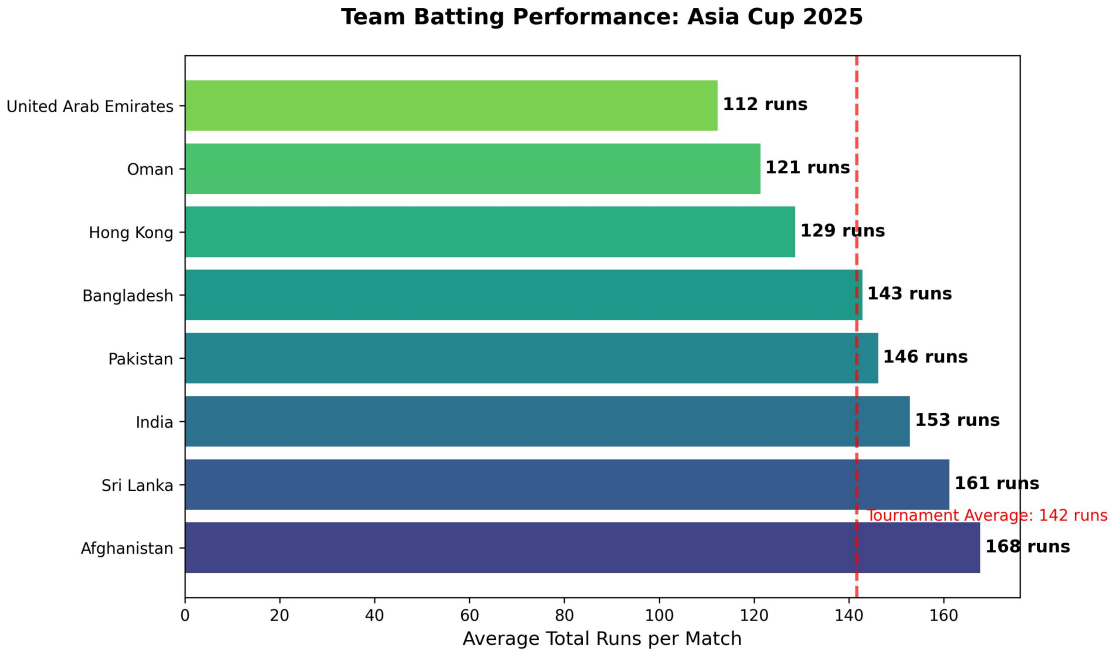

图2：2025年亚洲杯所有比赛的球队平均得分。

3.3 边界分析

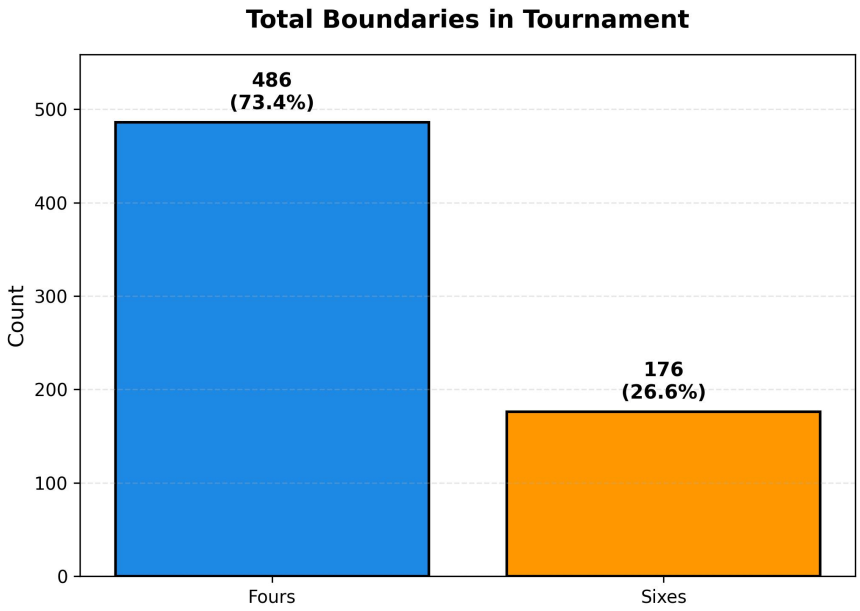

图 3：锦标赛期间记录的四分球和六分球分布。

4 数据可用性

数据集及其配套的分析代码已公开:

- **Zenodo 数据集：** `https://doi.org/10.5281/zenodo.17228056`
- **GitHub 仓库（探索性数据分析）：** `https://github.com/kousarraza/AsiaCup2025`

致谢

作者感谢**ESPNcricinfo** 和**亚洲板球理事会（ACC）**，它们是编制本数据集时使用的主要公开数据来源。

参考文献

[1]ESPNcricinfo。2025年亚洲杯比赛记分卡。 `https://www.espncricinfo.com/series/men-s-t20-asia-cup-2025-1496919`。[2] 亚洲板球理事会（ACC）。2025年亚洲杯官方锦标赛信息。 `https://www.asiancricket.org/`。[3] Raza, K.（2025）。2025年亚洲杯T20比赛级数据集。Zenodo。 `https://doi.org/10.5281/zenodo.17228056`。

[4] Raza, K.（2025）。2025年亚洲杯探索性数据分析代码。GitHub 仓库。
https://github.com/kousarraza/AsiaCup2025。