

# **Analyzing Umpire** Influence on Game Outcome

Are human umpires still the optimal baseball officiation method?

Project completed by Alex Nath in conjunction with Dr. Dahbura's Sports Analytics Research Group



### Context

Umpires play a critical role in baseball by enforcing rules and making judgment calls, particularly in pitch-calling, which significantly affects game flow and results. Given the subjective nature of these decisions, umpires are susceptible to cognitive biases that may consciously or unconsciously influence their calls.

#### **Initial Approach:**

- Admittedly dove into the project
- too eagerly, overlooking the
- overarching goal of the study.
  - Ran studies on broader game by game dataset

### **The Process**

#### **Overall home vs away bias:**



This ongoing study provides an initial examination into a few dimensions of umpire bias.

### Objectives

#### The goal:

- Investigate whether umpires are affected by traditionally suspected influences
  - Crowd
  - Garbage time
- Investigate how much of an effect the umpire's accuracy has on the

- Home vs away bias on individual umpire basis
- Home vs away bias on league wide basis
- Umpire accuracy vs game outcome

#### **Umpire home vs away bias:**





This study on home vs away bias revealed that umpires, in general, do not favor either team consistently.

#### **Accuracy vs Outcome:**



Left plot proves that absolute run differential is independent of umpire accuracy. Right chart shows the home team's advantage is constant regardless of change in the umpire's accuracy.



3est-fit Strike Zone and Associated Probability Surface for All Available Data in the 2008-2022 Seasons The 50% Probability Contour Defines the Strike Zone and is Shown with its Dimensions

#### outcome of a game



## Approach

**1.** Began using data from UmpireScoreCards.com with which I performed two major analyses:

- Individual umpire home vs away accuracy
- Overall umpire accuracy vs game outcome

2. Literature review and pivot in approach

- Umpires **consistently deviate** from the official zone in predictable ways
- Umpires are stricter on 2-strike counts, with significantly smaller strike zones on those pitches — primarily vertically (less top/bottom wiggle room)
- Pitchers like Bumgarner and Scherzer get **different calls** — Bumgarner benefits more on the edges
- Batters like Juan Soto get larger strike zones, possibly due to their stances or reputations. Aaron Judge's low zone is confirmed (bottom 3cm lower than average), though he gets smaller zones overall.



Horizontal Pitch Location (cm

My takeaways from the readings led to a shift in my approach, looking specifically at the biases emphasized in the papers





• Fastball and Changeup are high

from April to June 2024.

- Highest accuracy counts all have 2 strikes on
  - Batter is protecting and thus eliminates most of the close calls leaving only the obvious calls up to the umpire
- No strike counts have the lowest accuracy
  - Converse to the 2-strike count, batters are only swinging at their golden pitch while ahead. Therefore, the umpire is left to decide just about every other pitch.

Umpire Accuracy by Pitch Type (Apr-Jun 2024)

### Conclusion

While umpires generally maintain neutrality between teams, their accuracy varies significantly by pitch type and count — especially on two-strike pitches — revealing that certain game situations introduce consistent, predictable biases into officiating.

#### Literature:

Stockton, J., & Patt, E.-A. (2024). NOISY JUDGMENTS: A PROBABILITY SURFACE-BASED ANALYSIS OF UMPIRING VARIABILITY [Review of NOISY JUDGMENTS: A PROBABILITY SURFACE-BASED ANALYSIS OF UMPIRING VARIABILITY]. MIT Sloan Sports Analytics Conference. Jensen, P. (2009, June 28). Valuing the intentional walk. The Hardball Times. https://tht.fangraphs.com/thtlive/valuing-the-intentional-walk/

**3.** Accessed pitch by pitch Statcast data

• Umpire accuracy by count • Umpire accuracy by pitch



accuracy • Tend to move less across the strike zone making them easier to call • Slurve and Sinker are low accuracy • Late breaking pitches The dataset I used consisted of pitchspecific data for over 1,200 games



Scan the QR code to access a more extensive overview of the project. You will be taken to a loving document detailing findings throughout the study.