The Influence of Bat Speed & Swing Length on Fouling Off 2-Strike Pitches

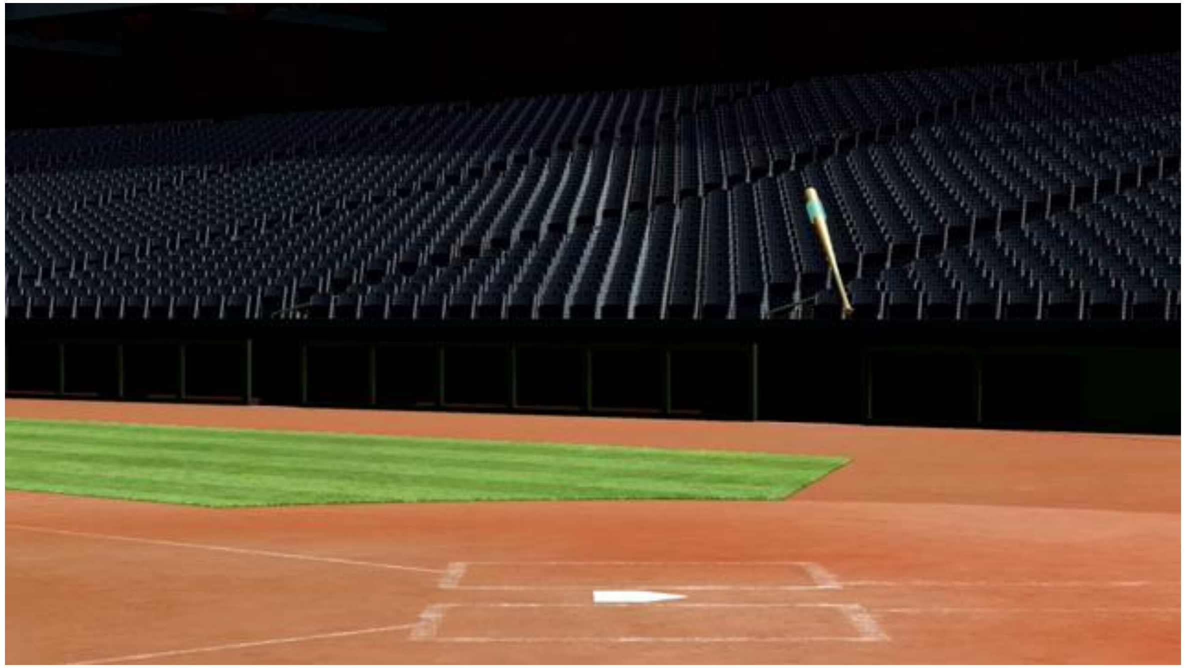
Ebenezer Olubayode[1], Dr. Daniel Larson[1]

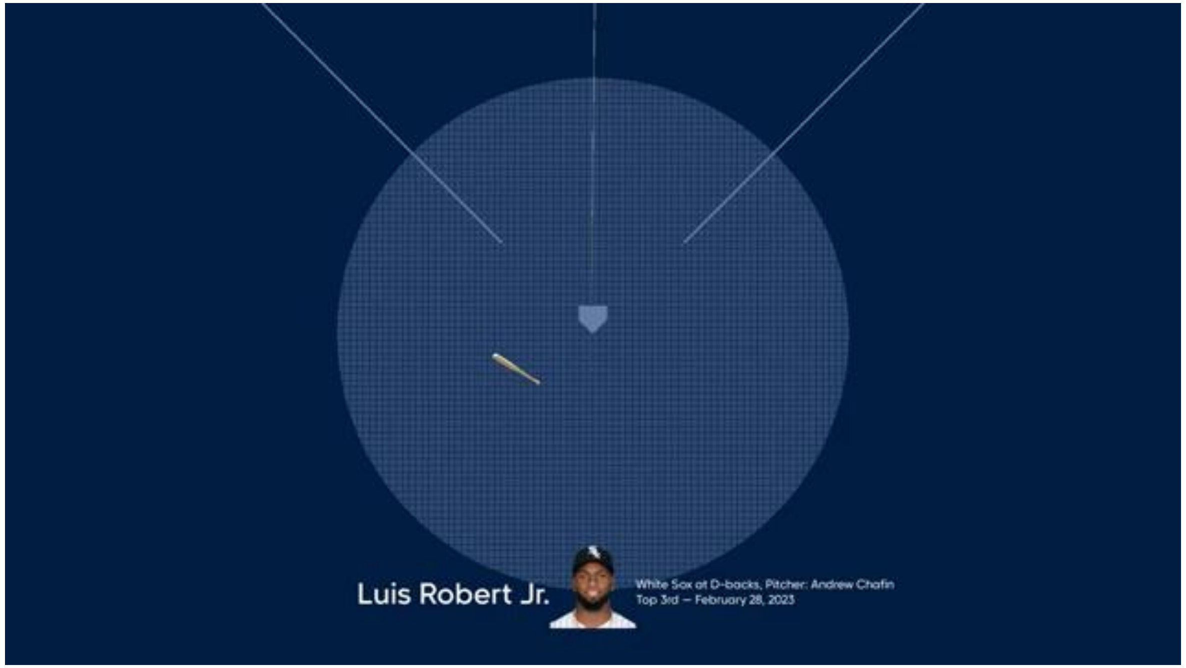
[1]Sports Data Analytics, Department of Health & Exercise Science, University of Oklahoma, USA

26th June 2025

What's New in MLB Tracking

- For the first time, we can measure how hitters move the bat through space:
  - *Bat speed, Swing path, attack angle*





MLB Statcast blog and Goldbeck's blog:
https://technology.mlblogs.com/introducing-statcast-2023-high-frame-rate-bat-and-biomechanics-tracking-3844890264a6
https://www.mlb.com/news/what-you-need-to-know-about-statcast-bat-tracking

Statcast Ball Tracking Data: Raising New Questions

- MLB's new bat-tracking shifts analysis from outcomes to the swing itself, sparking ***fresh questions*** about ***mechanics, timing, and plate discipline***.

- Until now, most of what we knew came from **what happened after contact**; *exit velocity, launch angle, hit outcome*.

  - ***How do different swings shape plate discipline?***
  - ***Can bat speed predict fouling ability in 2-strike counts? etc.***

- MLB clubs are exploring these questions to gain an edge

CONNECTICUT SPORTS ANALYTICS SYMPOSIUM (CSAS)

statds.org/events/csas2025/challenge.html

CSAS 2025

HOME
KEYNOTES
DATA CHALLENGE
SESSIONS
POSTERS
WORKSHOPS
PROGRAM
VENUE
REGISTRATION
COMMITTEES
RESOURCES
SPONSORS
HISTORY

- Graduate division
  - Reese, Boston University
  - Smart (Ebenezer Olubayode), University of Oklahoma

Overview

For this data challenge, your goal is to use new baseball data on bat speed and swing length to analyze some aspect of the pitcher/batter interaction. We provide pitch-level data from Baseball Savant for 346,250 Major League Baseball plate appearances from 4/2/2024 to 6/30/2024, including relevant Statcast data along with bat speed and swing length on pitches with a swing tracked. Data from the second half of the season will be added after the conclusion of the regular season. Your analysis should involve bat speed and swing length to study any topic related to the batter, pitcher, or batter-pitcher interaction during an at bat.

Since these data are new, there are a variety of topics that have not previously been studied. Below are a few example topics. However, we note that this list is far from exhaustive. Participants should feel free to study any aspect of the batter, pitcher, or batter-pitcher interaction that interests them, provided that bat speed and swing length are used in the analysis in some meaningful way.

- Batter Specific Plate Discipline. Can bat speed and swing length be used to measure some aspect of plate discipline?
  - Are these data useful in analyzing a batter's decision to swing at a pitch? For example, can faster swingers be more patient?
  - Are bat speed or swing length related to a hitter's ability to foul off 2-strike pitches, or protect the plate?
  - How is pitch location related to the above ideas? See https://baseballsavant.mlb.com/visuals/swing-take?playerId=545361 which may give you some ideas.
- Do players have different types of swings? Are these swing types related to situation (count, base runners, outs), pitcher, pitch location, pitch type, or anything else?
- Do pitchers modify batter behavior or do batters modify pitcher behavior (or neither! Or both!!):
  - Can pitchers "dictate" swings, or is it all the batter? Are swing lengths longer against certain pitchers? To what extent is that due to pitch velocity or spin rate, as opposed to other pitcher-specific factors?
  - Do pitchers modify behavior against batters with higher/lower bat speed (e.g. throw to different locations, throw different types of pitches, etc.)?

Data

The data is in this shared folder. We recently added a file that has data for the whole season (full regular season and playoffs), and has the column `arm_angle` included in the data along with some additional contextual variables and other information. The new columns were added to the right of the

https://statds.org/events/csas2025/challenge.html

Why Investigate Swing Mechanics in 2-Strike Fouling?

- Foul-rate behavior in two-strike situations isn't a ***fixed skill*** but ***variable and context-dependent***.
  - ***This hints at underlying factors driving those outcomes.***

- Foul rate rises in 2-strike counts but isn't consistent year-to-year.
  - ***Implies fouling isn't just talent-based but situational and mechanical***

- **Situational adjustment:** Hitters clearly adapt under pressure but how mechanically?
  - 🔍 ***Could bat speed and swing length be part of that adaptation?***

- **Mechanical dependency**: The article suggests fouling is tied to mechanics, not skill.
  - ✅ ***supporting the investigation into how specific swing traits influence fouling probability.***

Fox Sports. (2020, March 5). *Is fouling off pitches a skill?* https://www.foxsports.com/stories/mlb/is-fouling-off-pitches-a-skill

Why Investigate Swing Mechanics in 2-Strike Fouling?

Victor Martinez:

- He fouls off **19% more pitches with 2 strikes** than with **0–1 strikes**.
- **Low K% (6.6%)** and **High ISO (0.230),** he's not just fouling off, he's **protecting plate & producing power.**

Tommy Medica:

- **+16.0% IncreaseFoul%,** but **K% = 29.0%:** Despite increasing foul rate, **he still strikes out a lot**.
- Suggests his fouling doesn't even **reduce strikeouts** , possibly due to **mechanics**, **swing decisions**, or **pitch quality**.

- Not all succeed equally, and
- The variation across players hints at deeper factors (**like swing mechanics).**

FOX SPORTS

SCORES
WATCH
GOLD CUP
ODDS
FOX SUPER 6
STORIES
DAILY LISTS
SEARCH
SIGN IN

Take a look at 2014's leaderboard for the most extreme two-strike approaches. These are the (1,000-pitch minimum) leaders in foul rate in two-strike counts minus foul rate in one- and zero-strike counts ("IncreaseFoul%").

Batter	IncreaseFoul%	ISO	K%
Mike Moustakas	19.1%	0.149	14.8%
Victor Martinez	19.1%	0.230	6.6%
Dee Gordon	16.8%	0.089	16.5%
Tommy Medica	16.0%	0.175	29.0%
Jonny Gomes	15.9%	0.095	27.4%
Joe Mauer	15.8%	0.095	18.5%
Sam Fuld	15.7%	0.103	15.7%
Matt Carpenter	15.6%	0.103	15.7%
Stephen Vogt	15.5%	0.152	13.6%
Brett Gardner	15.3%	0.166	21.1%
Adam Eaton	15.2%	0.101	15.4%
Buster Posey	15.2%	0.179	11.4%
Jonathan Lucroy	15.0%	0.164	10.8%
Grady Sizemore	15.0%	0.121	19.9%
David DeJesus	14.5%	0.155	15.8%

Fox Sports. (2020, March 5). *Is fouling off pitches a skill?* https://www.foxsports.com/stories/mlb/is-fouling-off-pitches-a-skill

Why 2-Strike Fouls Matter?

- Fouling off 2-strike pitches helps batters *extend at-bats*, *tired pitchers*, *and increases on-base chances.*

- For pitchers, *preventing fouls* is key *to finishing counts and converting strikeouts*

- But how swing mechanics like bat speed and swing length affect this ability remains poorly understood.

Carleton, R. A. (2017, March 22). Baseball Therapy: The Secret Powers of the Foul Ball. Baseball Prospectus. https://www.baseballprospectus.com/news/article/31525/baseball-therapy-the-secret-powers-of-the-foul-ball/
Vock, D., & Vock, L. (2018). Estimating the effect of plate discipline using a causal inference framework: An application of the G-computation algorithm. Journal of Quantitative Analysis in Sports, 14(1). https://doi.org/10.1515/jqas-2016-0029

Research Objectives

- **Quantify** how **bat speed** and **swing length** influence the likelihood of fouling off **2-strike pitches**.
- **Evaluate** how **pitch characteristics** (vertical/horizontal movement) interact with swing mechanics to shape fouling outcomes

Fox Sports. (2020, March 5). *Is fouling off pitches a skill?* https://www.foxsports.com/stories/mlb/is-fouling-off-pitches-a-skill

Variable of Interests

Dependent Variable

Fouling off 2 Strike Likelihood (Y):

A binary outcome variable indicating whether a swing on a 2-strike pitch results in a foul ball (Y = 1) or not (Y = 0, which could be a miss, fair contact, or taken pitch) influenced by pitch type, location, and other factors.

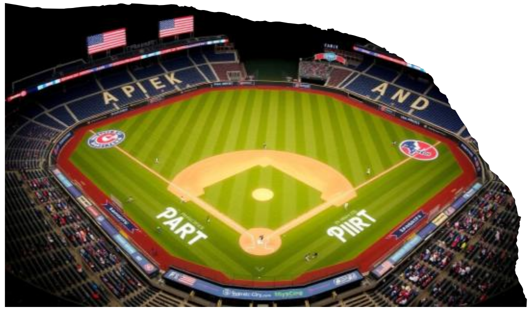
Independent Variables

Swing Length
The total (sum) distance in feet traveled of the head of the bat in X/Y/Z space, from start of tracking data until impact point.

Bat Speed
Bat speed is measured at the sweet-spot of the bat. Average bat speed is the average of the top 90% of a player's swings.

pfx_x

- Horizontal movement in feet from the catcher's perspective.



pfx_z

- Vertical movement in feet from the catcher's perpsective.

release_speed

- Pitch velocities

https://baseballsavant.mlb.com/csv-docs

Research Design and Data Collection

Dataset & Preprocessing

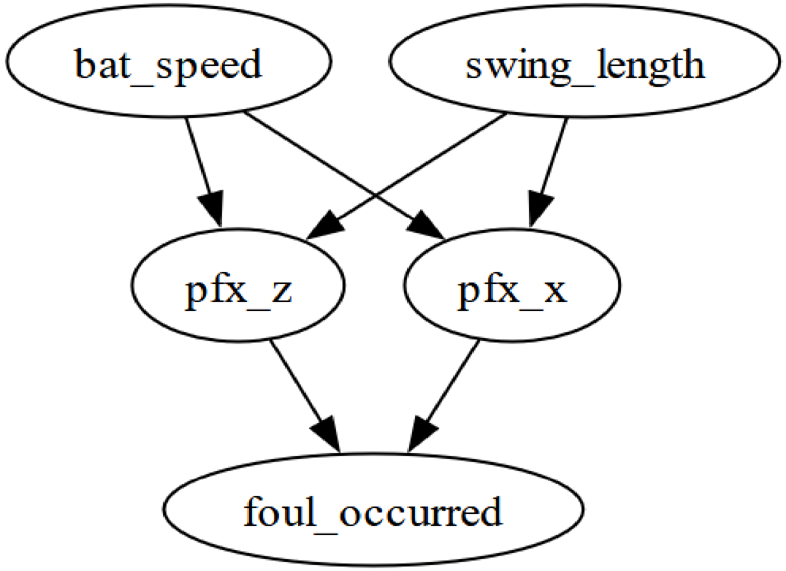
- Source: Statcast + MLB Savant (Half a season)
- Initial rows: *346,250 PAs → cleaning: 58,566*
    - 34,159 rows where fouling off 2-strikes didn't occur (foul_occurred = 0)
    - 24,407 rows where a fouling off 2-strikes occurred (foul_occurred = 1)

Study Framework:

- **Bayesian inference models** for causal analysis *(Megiddo, 2023*

Why Use Bayesian Modeling?

- Allows probabilistic reasoning
- Integrates prior knowledge
- Outputs full posterior distribution
- Used DAGs to establish **causal relationships** & adjust for confounding variables (Rohrer, 2018; Pearl, 2009).



Drawing the Causal Directed Acyclic Graph (DAG)

Bayesian Inference Workflow for Strike Outcomes

- Workflow follows a structured process to ensure **accurate causal estimates (Lundberg et al., 2021)**

1. Define Generative Model:

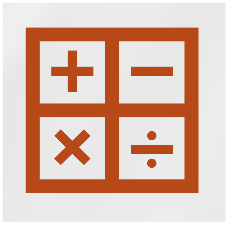
Identify all interested variables related to fouling off 2 strikes

Formulate the logistic regression model to represent the probability of a foul. **(Megiddo, 2023)**

2. Define Estimands:

Describe association between the factors **(Kahan et al., 2023).**

Foul outcomes given the input variables.



3. Design Estimator:

Choose appropriate priors for the coefficients to produce estimands.

Bayesian logistic regression applied to estimate **effect size and uncertainty (Gelman et al., 2008).**

.

Bayesian Inference Workflow for Strike Outcomes

4. Test Estimator:

- Generate synthetic data using the generative model to test the estimator
- Fit the Bayesian model to the synthetic data and evaluate the estimation accuracy **(Gogtay et al., 2021).**

5. Analyze and Summarize:

- Fit the Bayesian model to the actual data.
- Use (Markov Chain Monte Carlo) to sample from the posterior distributions in estimating the distribution of the model's parameters
- Interpret the impact of each predictor on the foul outcomes and computes **contrast means** for hypothesis testing (Wiens & Nilsson, 2016)
- **High-Density Intervals (HDI) ,** shows **where the true value is most likely to be (Hespanhol et al., 2019).**
  - *HDI is significant if it d*oes not include zero

Model Testing and Validation

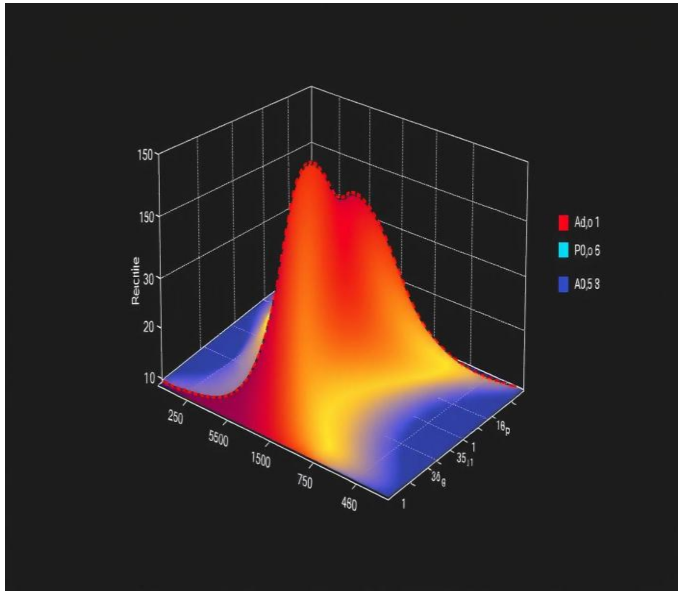
A rigorous testing was conducted *to ensure* ***model accuracy and reliability***, *(Didelez et al., 2010)*

📌 **Bayesian Priors and Their Role (Gelman et al., 2008):**

- **Prior distributions** represent **existing knowledge** before analyzing the data.

In this study, **Normal (0,10) priors was used** for model coefficients.

- **Why?** This prior assumes **no strong initial bias** while allowing flexibility in estimation.
- It acts as a **regularization tool**, preventing extreme estimates



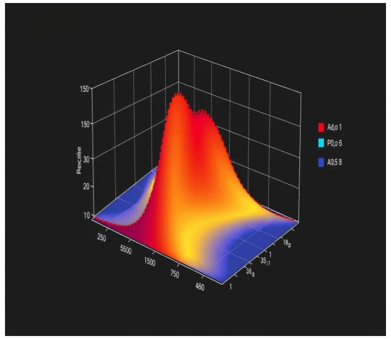
Model Testing and Validation

📌 **Convergence Diagnostics:**

Gelman-Rubin statistic (R-hat) checks chain stability (Hespanhol et al., 2019).

- If R-hat = 1, it means the MCMC chains have **stabilized**, ensuring reliable estimates.
- If R-hat > 1.1, it suggests the model **hasn't converged yet**, meaning more iterations are needed.

📌 **Effective Sample Size (ESS) evaluates MCMC sampling efficiency (Wiens & Nilsson, 2016).**



- A higher ESS (>1000) means **a stable and precise estimate** of parameters.
- A low ESS suggests **more iterations are needed** to reduce uncertainty

Results

Factor	Comparison	Contrasts	95% HDI	Interpretation
Swing Length	Short vs. Long	48% vs. 37% (-0.11)	[-0.137, -0.082] Significant effect, as HDI does not include zero	Shorter swing lengths increase fouling probability
Bat Speed	High vs. Low	41% vs. 44% (-0.035)	[-0,062, -0.008]	Higher bat speed slightly reduces fouling probability
Vertical Pitch Movement	Effect with Swing Length	0.50	[0.48, 0.53] High certainty of predicted value	Short swings are more effective against high pfx_z.
Bat Speed & Swing Length	High Bat Speed & Low Swing Length vs. High Bat Speed & High Swing Length	-0.105	[-0.145, -0.063]	Significant effect; Foul rate is **10.5% higher** with a short swing, even when bat speed is already high.
Bat Speed & Swing Length	Low Bat Speed & Low Swing Length vs. Low Bat Speed & High Swing Length.	-0.115	[-0.146, -0.083]	Significant effect; shorter swings increase fouling even with low bat speed

Main Findings

🔁 **Swing length is the dominant factor:**

- Short swings have a higher foul rate (48%) compared to long swings (37%), helping batters extend at-bats.

⚡ **Bat speed has a smaller effect than Swing length:**

- But low bat speed shows a slight increase in fouling (**44% vs 41%**) than high bat speed.

📈 **Vertical pitch movement (pfx_z) challenges discipline:**

- Higher vertical movement reduces fouling, but **short swings** help batters **adjust &** maintain contact.

🔗 **Interaction effects matter:**

- The **combo of short swings + high bat speed** yields the **highest fouling probability**, showing the **importance of mechanics over strength**.

Overall swing control matters more than raw power.

Skill or Talent? Implication for Player Development

Fox (2020) noted:

- **Fouling off 2-strike pitches isn’t just raw talent — it’s a skill**
- Not all players succeed equally in fouling off 2 strikes pitches… suggesting deeper factors like swing mechanics.

My findings confirm it:

  - Foul rates shift with swing length, bat speed, and pitch movement — They not consistent by player.

Takeaway:

- If it’s mechanical, it’s coachable.
  Hitters can train to adjust, shorten swings, and stay alive — especially in pressure counts.

This means:

- Success with 2 strikes comes more from **adjustment and control** than **natural talent.**

CONCLUSIONS

🧢 For Batters:

- **Short swings** are most effective tool for surviving 2-strike counts,
- especially against **pitches with ↑ vertical movement.**

- **Bat speed alone** has limited impact. Instead, **controlled swing mechanics**
- especially **reduced swing length drive plate protection**.

⚾ For Pitchers:

- To limit fouling, pitchers should **induce longer swings** by exploiting **pitch movement and location.**

🎯 **Targeting vertical movement** (e.g., riding fastballs) can help pitchers finish counts and increase strikeout chances.

Scan this

Github Code & References

THANK YOU

QUESTIONS ?

How Bayesian Markov Chain Monte Carlo Works

Step 1: Set a Prior (What We Believe Before Seeing Data)

🔮 **Prior Belief** → "Based on past seasons or coaching experience,

I think a low bat speed probably don't result in fouling off strikes by ~10%,

but could be wrong, so we open to a wide range of possibilities "

🧠 "That belief is our **prior distribution**, and it helps us frame expectations.

Example prior distribution: 🎯 N(–0.10, SD = 0.05)

←---|====|================|====|---→

-20% -10% 0% +5%

(Before looking at real data)

Step 2: Real Game Data

Actual Pitches → [bat_speed = 85.5mph, result = Fouling off 2 strikes]

[swing_lenght= 43inches, result = no No. fouling off 2 strikes]

... (real pitch data from games)

Step 3: MCMC Starts Sampling Repeating & Learning

We try different guesses

🧪 Guess 1 → Maybe a high bat speed results in foul by 5%?

🧪 Guess 2 → "Maybe a high bat speed don't result by 12%?"

🧪 Trial Guess 3 → "What if it's +2%?"

⬇️

📊 We test each guess by asking:

Does it match the data we saw?

Does it make sense based on our earlier belief?

⬇️

If yes → Keep it more often.

If no → Drop it or keep it rarely.

This back-&-forth helps us learn from both data and belief

MCMC is ***a trial-and-error engine that keeps the best guesses***.

This balance is the heart of Bayesian inference

Step 4: Posterior Distribution is Formed

Then we get a final result—**a posterior distribution.**

←---|====|======================|====|---→

-15% -10% -5% 0% +5%

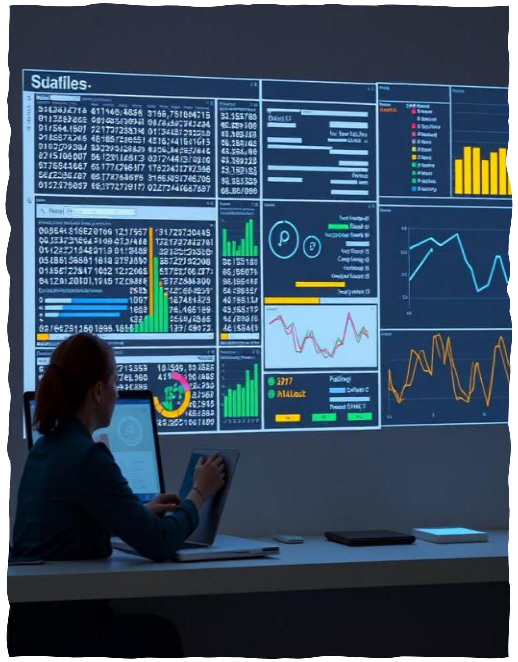
This shows where the true effect probably lies (Most likely effect: ~ -10%)

In this case, we now believe with 95% certainty that the real effect of bat speed on 2-strike fouling is between -15% and -5%.

Data Sampling Justification

- Selected to **balance computational feasibility** with **posterior estimation reliability**.
- Literature supports sample sizes of **2,000–5,000 observations** for accurate posterior estimation using efficient samplers like **NUTS.** *Gelman et al. (2013); Kruschke (2014); Vehtari et al. (2017)*.
- Larger datasets may be used in **future predictive modeling** where generalizability is prioritized.

Statistical Analysis – Bayesian Modeling



📌 **Confounding & Mediation:** DAGs **establish causal relationships** (Rohrer, 2018; Pearl, 2009). Identifying **covariates, mediators, and biases**, ensuring **valid causal inferences.**

📌 **Causal structure:** Innings treated as confounders; movement modeled as mediator.

📌 **Estimation method:**

Markov Chain Monte Carlo (MCMC) sampling estimates posterior distributions **(Chen, 2024).**

- Instead of calculating a **single best estimate**, MCMC **samples multiple possibilities** ensure that **uncertainty** in predictions are captured.

Assumptions

📌 **Weakly Informative Priors:** assume predictor effects follow **Normal (0,10) priors**, ensuring flexibility while preventing extreme values.

Rover et al. (2020) emphasize the benefits of **weakly informative priors in hierarchical models**, helping avoid overfitting.

📌 **Minimal External Influence:** Factors like **weather conditions** may affect pitch resulting in strike but are assumed (**negligible effects).**

📌 Assumed that batter-specific behavior could bias strike outcomes, so fixed effects were included to control for individual batting tendencies.

📌 Assumed that pitchers exhibit natural variability in style, so random effects were used to capture individual differences without overfitting.