



# Balancing Performance and Longevity: Modeling Injury Risk in MLB Pitchers

STATS140XP

Group 5: Oscar Su, John Tan, Hannie Xie, Patrick Xue, Erica Yee





# Table of contents

## 01 Introduction

Motivation, Research Questions

## 03 Research Question #2

Methods, Results, Interpretation

## 02 Research Question #1

Methods, Results, Interpretation

## 04 Conclusion + Recommendations

Assumptions, Takeaways





01

# Introduction





# Project Poster

**UCLA**

## Balancing Performance and Longevity: Modeling Injury Risk in MLB Pitcher

Oscar Su<sup>1</sup>, John Tan<sup>1</sup>, Hannie Xie<sup>1</sup>, Patrick Xue<sup>1</sup>, Erica Yee<sup>1</sup>  
1University of California, Los Angeles

### Introduction

Major League Baseball (MLB) teams invest substantial resources in developing young pitchers, yet early-career pitchers are workflowed, and durability vary dramatically across players. As a result, identifying reliable early indicators of long-term success and injury risk is critical for team success.

This project investigates whether a pitcher's **first 10 to 15 MLB seasons** provide meaningful predictive signals for:

- Career longevity** – which early-career traits distinguish pitchers who remain in MLB from those who don't quickly leave?
- Injury risk** – which early-career metrics, such as strikeout rate or usage patterns, predict future injury or restricted performance thresholds?

By integrating early-career physical attributes, pitch characteristics, and game log patterns, our goal is to identify the factors that most strongly shape both career longevity and MLB career viability.

### Data and Study Methods

The dataset was obtained from [MLB.com](#). Primary variables include player statistics, game logs, and pitch data. We also extract pitch-by-pitch information including pitch velocity, pitchtype, release point, and game outcome. Our dataset aggregates these pitch-level measures over season summaries.

**Career Length**

- Compare long vs. short MLB career lengths using univariate t-tests
- Fit a multiple linear regression model to quantify which early-career variables predict total career length, using backward selection to remove irrelevant or non-informative predictors

**Injury Risk & Performance**

- Build a logistic regression model to estimate the probability of being placed on the injury list within 14 days, based on early-career pitch-level features and workload
- Use Poisson regression models to model discrete outcomes (e.g., BBs, HRs, outs) and assess how counts of fastball, breaking balls, and off-speed pitches influence each outcome

### Research Question 1: Which Variables Contribute to a Long Career?

**Background**

We grouped the metrics into four key categories:

- Physicality**: Height, weight, total days pitched
- Command**: Strikeout rate, walk rate, ERA
- Workload**: Velocities of fastball, breaking, off-speed pitches and average pitches per game
- Style of Pitch**: Proportions of fastball, breaking, and off-speed pitches

To allow for a more fine-grained and accurate comparison between players at similar career stages, we restrict our analysis to pitchers who:

- Debut between 2010–2015, giving them the opportunity to reach 5+ MLB seasons
- Have at least 10 seasons of data available for performance data

**Figure 1: Distribution of total MLB career length (in years) by pitchers**

**Observations**

The highest concentration of pitchers falls in the 3–10 year range. This steep early peak reflects how quality players rise in and out of the top tier. After age 10, the number of players who survive beyond this early career window declines considerably, suggesting that remaining competitive past these milestones acts as a natural filter for skill, command, health, and durability.

Only a small fraction of pitchers achieve 10+ years.

**Interpretations**

This finding suggests that predicting career longevity is an important problem. With such a large proportion of pitchers exiting MLB within their first few seasons, identifying the early indicators that can distinguish between pitchers who go on to have long careers and those who do not is critical for team success. For regulators, using these insights can improve decisions related to player development, investment planning, and injury prevention.

### 1. Univariate Analysis of 10-Year

We test whether pitchers with short careers ( $< 5$  years) differ significantly from those with long careers ( $\geq 5$  years).

**Model Fit:** All early-career indicators show the same short and long career pitchers.

**Table 1: Least Absolute Shrinkage and Selection Operator (LASSO) Results**

Variable	Estimate	Std. Error	t Statistic	p-value
Total Days	0.734	23.124	254.48	< 0.001
Days Worked	0.734	23.124	254.48	< 0.001
BB/P	0.637	4.74	13.42	< 0.001
Avg Fastball Velocity	50.500	0.000	250.00	< 0.001
Avg Offspeed Velocity	60.617	83.000	0.73	0.464
Avg Breaking Velocity	50.500	0.000	250.00	< 0.001
HRs	0.634	0.251	2.52	0.012
Walks	0.634	0.251	2.52	0.012

**Table 2: Results**

The only statistically significant predictor of career length ( $p < 4.03 \times 10^{-10}$ ) was:

- BB/P**: Pitchers who walk fewer hitters last longer. Good command is a key indicator of sustained MLB success.
- HRs**: Higher fastball velocity predicts career longevity. A faster fastball means better mechanics and greater durability.
- Avg Offspeed Velocity**: Higher offspeed velocity also correlates with longer careers, suggesting better mechanics and greater durability.

This suggests that pitch velocity (both fastball and offspeed) and Walks per nine innings may be the strongest early-career signs of whether a pitcher will remain competitive in MLB.

### 2. Multiple Linear Regression

A backward stepwise selection process finds three key predictors of career length:

**Model Fit:**  $R^2 = 0.088$ , Adjusted  $R^2 = 0.085$ , Standard Error: 0.2 - 2.700, F-statistic: 1.67, p-value: < 0.05

**Variables:** Total Days = 1.13, BB/P = 1.13, Avg Fastball Velocity = 1.09 (the multicollinearity issues)

**Figure 2a: Residuals vs Fitted Values**

**Figure 2b: Q-Q Residuals**

**Figure 2c: Scale Location**

**Figure 2d: Residual vs Leverage**

**Research Question 2(a): How do different pitch types affect a pitcher's risk of injury?**

**Background**

Injury probability indicates such as strikeouts (K), walks (W), home runs (HR), and outs are discrete count variables, and highly right-skewed, as shown in the distributions below. Because of this, we model these outcomes using Poisson regression.

**Figure 3: Distribution of pitch types**

**Poisson Regression for Performance**

**Figure 4: Residuals vs Fitted Values**

**Figure 5: Q-Q Residuals**

**Figure 6: Scale Location**

**Figure 7: Residual vs Leverage**

**Research Question 2(b): How do different pitch types affect a pitcher's performance?**

**Background**

Key performance indicators such as strikeouts (K), walks (W), home runs (HR), and outs are discrete count variables, and highly right-skewed, as shown in the distributions below. Because of this, we model these outcomes using Poisson regression.

**Effect of Pitch Counts on K**

Variable	Estimate	Std. Error	t Statistic	p-value
Total Fastball	-0.035	0.005	-7.00	< 0.001
Total Break	-0.035	0.005	-7.00	< 0.001
Total Offspeed	-0.035	0.005	-7.00	< 0.001
Fastball Count	-0.035	0.005	-7.00	< 0.001
Break Count	-0.035	0.005	-7.00	< 0.001
Offspeed Count	-0.035	0.005	-7.00	< 0.001

**Effect of Pitch Counts on W**

Variable	Estimate	Std. Error	t Statistic	p-value
Total Fastball	0.035	0.005	7.00	< 0.001
Total Break	0.035	0.005	7.00	< 0.001
Total Offspeed	0.035	0.005	7.00	< 0.001
Fastball Count	0.035	0.005	7.00	< 0.001
Break Count	0.035	0.005	7.00	< 0.001
Offspeed Count	0.035	0.005	7.00	< 0.001

**Effect of Pitch Counts on HR**

Variable	Estimate	Std. Error	t Statistic	p-value
Total Fastball	0.035	0.005	7.00	< 0.001
Total Break	0.035	0.005	7.00	< 0.001
Total Offspeed	0.035	0.005	7.00	< 0.001
Fastball Count	0.035	0.005	7.00	< 0.001
Break Count	0.035	0.005	7.00	< 0.001
Offspeed Count	0.035	0.005	7.00	< 0.001

**Effect of Pitch Counts on Out**

Variable	Estimate	Std. Error	t Statistic	p-value
Total Fastball	0.035	0.005	7.00	< 0.001
Total Break	0.035	0.005	7.00	< 0.001
Total Offspeed	0.035	0.005	7.00	< 0.001
Fastball Count	0.035	0.005	7.00	< 0.001
Break Count	0.035	0.005	7.00	< 0.001
Offspeed Count	0.035	0.005	7.00	< 0.001

**Model Fit:** Residual deviance = 254.36, AIC = 254.3

**Figure 8: ROC Curve**

**Figure 9: Area Under Curve (AUC)**

**Table 3: Summary of Poisson Coefficient Interpretation for Each Outcome**

Outcome	Max Associated Pitch Type	Report. t Statistic	Report. p-value
Strikeout	Breaking	1.2290	$< 0.001$
Walk	Fastball	0.0015	$< 0.001$
Home Run	Fastball	0.0015	$< 0.001$
Out	Fastball	0.0015	$< 0.001$

### Conclusion

**Key Findings**

- Longevity (Q1):** Longer careers are associated with lower BB/P, higher fastball velocity, and earlier career workload.
- Injury Risk (Q2):** Only early fastball volume predicts injury frequency (breaking counts are not significant). Fastballs carry the highest risk of injury, while off-speed pitches carry the lowest risk.
- Performance (Q3):** Strikeouts are the best indicator of performance, followed by walks, then home runs, then outs. Breaking balls are better than fastballs, fastball carry higher risk of injury than off-speed pitches.

**Limitations**

- Sensitivity to injuries:** Some injuries show negative or irregular sensitivity values due to pre-MLB practice or other sources, which may not reflect early-career status.
- Carrying of career length:** Data ends in 2020, so pitchers labeled as "short career" may still be active or returning to the league.
- Injury list ambiguity:** Injury List (IL) placement is not always strictly medical; teams may use it as a strategic move, such as to prevent a pitcher from being traded.
- Unobserved confounding:** We lack biomechanics, training load, medical history, and conditioning data of all pitchers, which may bias our results.

### Recommendations

**Injury Prevention**

Teams should closely monitor early-career workload, as sustained increases of high fastball usage negatively increase injury risk. Additionally, pitchers should be encouraged to develop more efficient delivery techniques before becoming harmful. Recovery should be tailored to prevent pitch volume rather than relying solely on rest day counts, which often fail to reflect fatigue.

**Payer Development**

MLB teams should invest in agent analysis, developing command, since lower walk rates consistently predict longer careers. Improving breaking ball quality offers a great upside for generating strikeouts, while refining off-speed pitches is also highly recommended by home runs. Together, these adjustments help pitchers balance efficiency with long-term durability.

**Scouting and Evaluation**

MLB scouts should focus on projects who create velocity with stable command, as this pairing strongly signals future longevity. Early pitch velocity tends to be more consistent than breaking ball velocity, and thus more reliable. Combining pitch-level data with biomechanical assessments provides a more complete picture of a pitcher's long-term potential.



# Motivation



**Performance ↔ health:** Research shows higher pitching velocity and heavier workloads are strongly associated with higher injury risk in MLB pitchers.

Increased velocity raises stress on the shoulder/elbow → **UCL injuries** that may require Tommy John surgery (Krause).

**Why it matters:** Recovery commonly keeps pitchers out for ~17 months, which can stall development for younger players and end careers for veterans

Source: Krause, Matthew. "Matthew Krause." Professional Baseball Strength, 5 Jan. 2024, [pbsccs.org/the-effects-of-high-velocity-on-players-health-safety-and-performance/](https://pbsccs.org/the-effects-of-high-velocity-on-players-health-safety-and-performance/).

# Research Questions

**Question 1:** Which variables contribute the most to a long career length?

**Question 2:** Are there specific pitches that lead to

- (a) Higher chance of injury
- (b) Better performance





# 02

RQ1: Which variables contribute the most to a long career length?





# Variables

Type	Metric	Description
Physicality	Height	Height of player in inches. We take the first recorded height when the player enters the dataset (assume no drastic change in height)
	Weight	Weight of player in lbs. We average their weight across the two seasons.
	Total_Days_Injured	Total cumulative number of days injured
Core Performance	K_per_9	Number of strikeouts per 9 innings (a measure of performance for a pitcher; outs are good)
	BB_per_9	Number of walks per 9 innings (a measure of blunders for a pitcher; walks are bad and load the bases)
	HR_per_9	Number of home runs given per 9 innings (a measure of blunders for a pitcher; home runs are really bad)





# Variables

Workload	Avg_Fastball_Overall	Average speed of a fastball over their first 2 seasons
	Avg_Breaking_Overall	Average speed of a breaking ball over their first 2 seasons
	Avg_Offspeed_Overall	Average speed of an offspeed ball over their first 2 seasons
	Avg_Pitches_Per_Game	Average number of pitches per game over their first 2 seasons
Style of Pitch	Prop_Fastball	Average proportion of fastballs thrown by a player over their first 2 seasons
	Prop_Breaking	Average proportion of breaking balls thrown by a player over their first 2 seasons
	Prop_Offspeed	Average proportion of offspeed balls thrown by a player over their first 2 seasons



# RQ1 Methods: t-tests + stepwise linear regression



- Sample: MLB pitchers who debuted 2010–2015
- Predictors window: summarize each pitcher's first 2 seasons (early-career metrics + pitch mix)
- Outcome label: career length > 5 years = long, otherwise short (to reduce survivorship bias)



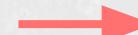
Univariate  
t-tests



Multiple  
Linear  
Regression



Backward  
Stepwise  
Selection





# RQ1 t-test Results

Variable <chr>	P_Value <dbl>	Mean_Short_Career <dbl>	Mean_Long_Career <dbl>	Significant <chr>
1 Height	0.58603	74.279	74.476	NO
2 Weight	0.73358	213.162	214.397	NO
3 Total_Days_Injured	0.56250	4.191	5.952	NO
4 K_per_9	0.28747	7.001	7.566	NO
5 BB_per_9	0.01740	4.785	3.624	YES
6 HR_per_9	0.09470	1.381	0.990	NO
7 Avg_Fastball_Overall	0.00095	90.983	92.426	YES
8 Avg_Breaking_Overall	0.13390	80.489	81.409	NO
9 Avg_Offspeed_Overall	0.01660	83.003	84.334	YES
10 Avg_Pitches_Per_Game	0.05437	33.760	42.903	NO

Variable <chr>	P_Value <dbl>	Mean_Short_Career <dbl>	Mean_Long_Career <dbl>	Significant <chr>
11 Prop_Fastball	0.25529	0.570	0.590	NO
12 Prop_Breaking	0.81394	0.205	0.201	NO
13 Prop_Offspeed	0.63333	0.089	0.082	NO

- **BB\_per\_9:** Players who give lower walks per 9 on average have longer careers.

- **Avg\_Fastball\_Overall:** Important to have a faster than average fastball.

- **Avg\_Offspeed\_Overall:** Also important to have a faster offspeed ball.



# RQ1 MLR and Backward Stepwise Selection

Career Length =  $-44.2 + 0.00684(\text{Total Outs}) - 0.244(\text{BB_per_9}) + 0.533(\text{Avg Fastball Velocity})$

Call:

```
lm(formula = Career_Length ~ Total_Outs + BB_per_9 + Avg_Fastball_Overall,  
   data = model_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.1094	-2.2065	0.0238	2.1264	5.3589

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-44.200341	9.096232	-4.859	3.79e-06 ***
Total_Outs	0.006840	0.001387	4.930	2.82e-06 ***
BB_per_9	-0.244467	0.115651	-2.114	0.0367 *
Avg_Fastball_Overall	0.533436	0.099077	5.384	3.95e-07 ***

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.704 on 114 degrees of freedom

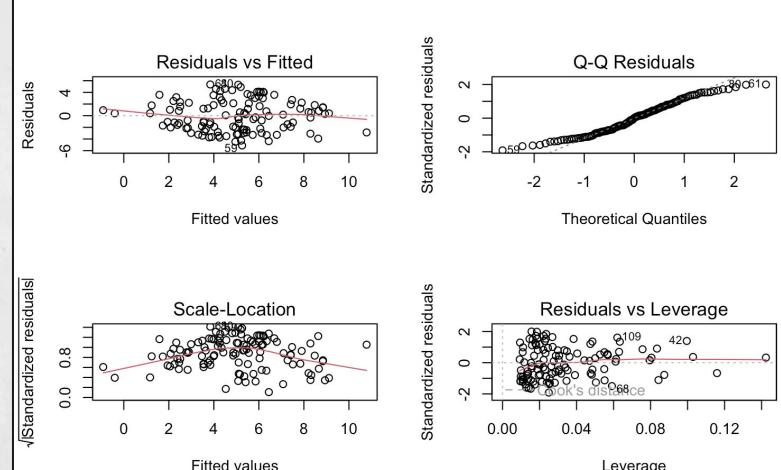
Multiple R-squared: 0.3829, Adjusted R-squared: 0.3667

F-statistic: 23.58 on 3 and 114 DF, p-value: 6.036e-12

vif(backward\_model)

````

|  | Total_Outs | BB_per_9 | Avg_Fastball_Overall |
|--|------------|----------|----------------------|
|  | 1.149446   | 1.149504 | 1.000127             |



No multicollinearity, OK diagnostic plots

# RQ1 Interpretation: Key early-career predictors of longevity (debut 2010–2015)



Command matters: Long-career pitchers issue fewer walks (BB/9) early on.

Velocity matters: Long-career pitchers have higher average fastball velocity (and offspeed velocity shows up in group differences too).

Early dominance matters: In the final OLS model, the strongest predictors retained are Total Outs (+), BB/9 (-), Avg Fastball Velocity (+)

Coaches should prioritize developing and investing in young pitchers who, early in their careers, record more outs, issue fewer walks, and sustain higher fastball velocity, since these factors are associated with greater long-term value and a higher likelihood of career longevity.





# 03

RQ2: How do different  
pitch types affect pitcher's  
(a) risk of injury  
(b) performance?





## RQ2(a) How do different pitch types affect a pitcher's risk of injury?

- Focus on new players in their **debut** season
- Predictors: **3-game cumulative frequency** of different **pitch types** (fastball / breaking / offspeed) and **number of rest days**
- Outcome: **potential injuries** (placed on the injury list within the next 14 days)



Logistic Regression  
On Injury





## RQ2(a) Analysis: Logistic Regression on Injury

```
Call:  
glm(formula = fml, family = "binomial", data = new_players_pitch_count)
```

Coefficients:

|                | Estimate  | Std. Error | z value | Pr(> z )    |
|----------------|-----------|------------|---------|-------------|
| (Intercept)    | -5.764673 | 0.535299   | -10.769 | < 2e-16 *** |
| roll3_fastball | 0.029684  | 0.009787   | 3.033   | 0.00242 **  |
| roll3_breaking | -0.020956 | 0.023737   | -0.883  | 0.37731     |
| roll3_offspeed | -0.031495 | 0.033183   | -0.949  | 0.34255     |
| rest_days      | 0.018842  | 0.018428   | 1.022   | 0.30655     |

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 265.19 on 2709 degrees of freedom  
Residual deviance: 254.30 on 2705 degrees of freedom  
AIC: 264.3
```

Number of Fisher Scoring iterations: 8

roll3\_fastball\_count is positive and significant ( $\beta \approx 0.0297$ ,  $p \approx 0.0024$ )





## RQ2(a) Analysis Interpretation

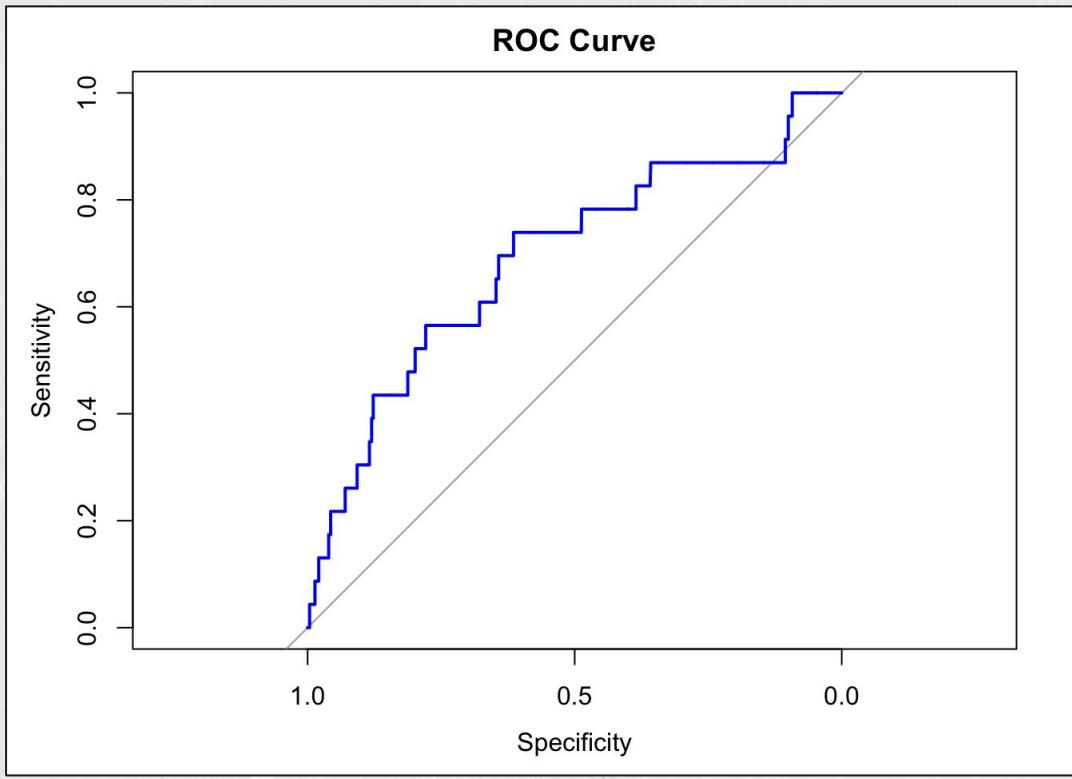


$$\text{Odds Ratio} = \exp(0.0297) \approx 1.03$$

- Odds interpretation: each extra fastball in the last 3 games → ~3% higher odds of injury
- Breaking/offspeed: not significant in this short-term window
- Rest days: not significant (may be too coarse / young pitchers recover quickly)



# RQ2(a) Analysis #1: Logistic Regression Model Quality

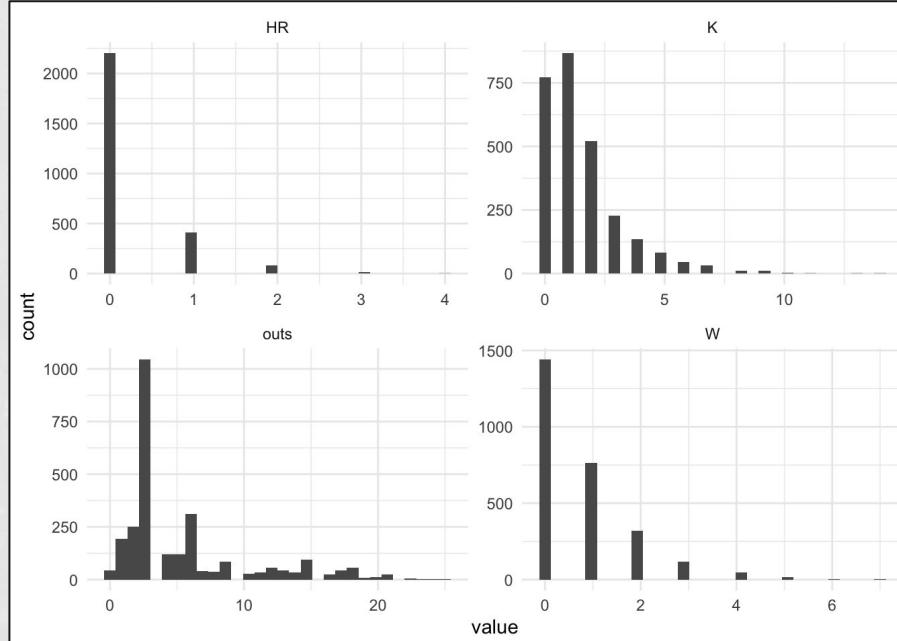


AUC  $\approx 0.689$





## RQ2(b) How do different pitch types affect a pitcher's performance?



**Why Poisson:** KPIs (outs, walks, HR, strikeouts) are discrete + right-skewed

**Model form:**  $\log(E[\text{outcome}]) = \beta_0 + \beta_1 \cdot \text{fastball} + \beta_2 \cdot \text{breaking} + \beta_3 \cdot \text{offspeed}$

larger coefficient  $\Rightarrow$  stronger association with that outcome

Poisson Regression on Performance



# RQ2(b) Analysis #2: Poisson Regression on Performance



Effect of Pitch Counts on K

| Variable       | Estimate | Std. Error | p-value                |
|----------------|----------|------------|------------------------|
| Intercept      | -0.4185  | 0.0282     | $< 2 \times 10^{-16}$  |
| Fastball Count | 0.0189   | 0.0009346  | $< 2 \times 10^{-16}$  |
| Breaking Count | 0.0284   | 0.0017070  | $< 2 \times 10^{-16}$  |
| Offspeed Count | 0.0195   | 0.0024030  | $5.61 \times 10^{-16}$ |

Effect of Pitch Counts on HRs

| Variable       | Estimate | Std. Error | p-value                |
|----------------|----------|------------|------------------------|
| Intercept      | -2.5381  | 0.0782     | $< 2 \times 10^{-16}$  |
| Fastball Count | 0.0190   | 0.002372   | $1.11 \times 10^{-15}$ |
| Breaking Count | 0.0325   | 0.004235   | $1.59 \times 10^{-14}$ |
| Offspeed Count | 0.0386   | 0.005669   | $9.26 \times 10^{-12}$ |

Effect of Pitch Counts on Walks

| Variable       | Estimate | Std. Error | p-value                |
|----------------|----------|------------|------------------------|
| Intercept      | -1.2290  | 0.0415     | $< 2 \times 10^{-16}$  |
| Fastball Count | 0.0251   | 0.001324   | $< 2 \times 10^{-16}$  |
| Breaking Count | 0.0210   | 0.002574   | $3.23 \times 10^{-16}$ |
| Offspeed Count | 0.0137   | 0.003470   | $7.78 \times 10^{-5}$  |

Effect of Pitch Counts on Outs

| Variable       | Estimate | Std. Error | p-value               |
|----------------|----------|------------|-----------------------|
| Intercept      | 0.7911   | 0.0152     | $< 2 \times 10^{-16}$ |
| Fastball Count | 0.0217   | 0.000493   | $< 2 \times 10^{-16}$ |
| Breaking Count | 0.0229   | 0.000936   | $< 2 \times 10^{-16}$ |
| Offspeed Count | 0.0243   | 0.001232   | $< 2 \times 10^{-16}$ |

★ All coefficients are positive and statistically significant, but their magnitudes vary by outcome



## RQ2(b) Analysis #2: Interpretation



| Outcome        | Pitch type with strongest association | Direction                                    |
|----------------|---------------------------------------|----------------------------------------------|
| Strikeouts (K) | Breaking                              | ↑ more breaking →<br>↑ expected Ks           |
| Walks (W)      | Fastball                              | ↑ more fastballs →<br>↑ expected BBs (walks) |
| Home Runs (HR) | Offspeed                              | ↑ more offspeed →<br>↑ expected HRs          |
| Outs           | About the same across all             | similar association<br>across pitch types    |



# RQ2 Analysis #2: Poisson Regression on Performance

```
summary(final_model_results$K$Best_Model_Object)
```

Call:  
glm(formula = fml, family = poisson, data = data)

Coefficients:

|             | Estimate   | Std. Error | z value | Pr(> z )     |
|-------------|------------|------------|---------|--------------|
| (Intercept) | -0.4184550 | 0.0282053  | -14.836 | < 2e-16 ***  |
| fastball    | 0.0188992  | 0.0009346  | 20.221  | < 2e-16 ***  |
| breaking    | 0.0284338  | 0.0017070  | 16.657  | < 2e-16 ***  |
| offspeed    | 0.0194577  | 0.0024030  | 8.097   | 5.61e-16 *** |

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 4753.1 on 27 summary(final\_model\_results\$outs\$Best\_Model\_Object)  
Residual deviance: 2765.9 on 27  
AIC: 7675.5

Number of Fisher Scoring iterations: Call:  
glm(formula = fml, family = poisson, data = data)

Coefficients:

|             | Estimate  | Std. Error | z value | Pr(> z )   |
|-------------|-----------|------------|---------|------------|
| (Intercept) | 0.7911495 | 0.0152265  | 51.96   | <2e-16 *** |
| fastball    | 0.0217320 | 0.0004930  | 44.09   | <2e-16 *** |
| breaking    | 0.0228794 | 0.0009356  | 24.45   | <2e-16 *** |
| offspeed    | 0.0243477 | 0.0012328  | 19.75   | <2e-16 *** |

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 9543.7 on 2709 degrees of freedom  
Residual deviance: 1693.1 on 2706 degrees of freedom  
AIC: 10574

Number of Fisher Scoring iterations: 4

```
summary(final_model_results$W$Best_Model_Object)
```

Call:  
glm(formula = fml, family = poisson, data = data)

Coefficients:

|             | Estimate  | Std. Error | z value | Pr(> z )     |
|-------------|-----------|------------|---------|--------------|
| (Intercept) | -1.229033 | 0.041485   | -29.626 | < 2e-16 ***  |
| fastball    | 0.025095  | 0.001324   | 18.955  | < 2e-16 ***  |
| breaking    | 0.021017  | 0.002574   | 8.164   | 3.23e-16 *** |
| offspeed    | 0.013711  | 0.003470   | 3.951   | 7.78e-05 *** |

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 3675.3 on 2709 degrees  
Residual deviance: 2582.6 on 2706 degrees  
AIC: 5526.6

Number of Fisher Scoring iterations: 5

Call:

glm(formula = fml, family = poisson, data = data)

Coefficients:

|             | Estimate  | Std. Error | z value | Pr(> z )     |
|-------------|-----------|------------|---------|--------------|
| (Intercept) | -2.538112 | 0.078153   | -32.476 | < 2e-16 ***  |
| fastball    | 0.019012  | 0.002372   | 8.014   | 1.11e-15 *** |
| breaking    | 0.032527  | 0.004235   | 7.680   | 1.59e-14 *** |
| offspeed    | 0.038646  | 0.005669   | 6.817   | 9.26e-12 *** |

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 2177.3 on 2709 degrees of freedom  
Residual deviance: 1757.7 on 2706 degrees of freedom  
AIC: 2847.1

Number of Fisher Scoring iterations: 6





04

# Conclusion



# Limitations



## Data quirks:

- Some players have negative seniority years (not official MLB debut; practice/offseason games), which may inflate performance.

## Incomplete window:

- Missing games, and the dataset cuts off at 2020 → careers may continue past the data (or players may disappear due to breaks/injury).

## Injury label isn't purely medical:

- “Injury list” can reflect roster strategy (resting, demotion), not just true injury incidence.

## Unobserved confounding:

- Lack player-level health/lifestyle/conditioning info that could affect both injuries and longevity.





# Recommendations



## Injury Prevention

- Closely monitor rolling fastball workload
- Recovery plans should be tailored to recent pitch volume rather than relying solely on rest-day counts

## Player Development

- Emphasize sharpening command as lower walk rates consistently predict longer careers
- Improve breaking-ball quality → greatest upside for generating strikeouts, refining offspeed pitches → reduce susceptibility to home runs
- Help pitchers balance effectiveness with long-term durability.

## Scouting and Evaluation

- Prioritize identifying prospects with strong fastball quality as this pitch is the foundation of pitching success



# Thanks!

**CREDITS:** This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Fleepik](#)

Please keep this slide for attribution