Balancing Performance and Longevity: Modeling Injury Risk in MLB Pitcher

Oscar Su [1] John Tan [1] Hannie Xie [1] Patrick Xue [1] Erica Yee [1]

[1]University of California, Los Angeles

Introduction

Major League Baseball (MLB) teams invest substantial resources in developing young pitchers, yet early-career performance, workload, and durability vary dramatically across players. As a result, identifying reliable early indicators of long-term success remains a central challenge for player development, roster planning, and injury prevention.

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

1. **Career longevity** — which early-career traits distinguish pitchers who remain in MLB from those who exit quickly?
2. **Injury and performance outcomes** — do specific pitch types or usage patterns elevate injury risk or reinforce performance strengths?

By integrating early-career physical attributes, pitch characteristics, and workload patterns, our goal is to identify the factors that most strongly shape both **immediate effectiveness** and **sustained MLB career viability**.

Data and Study Methods

The dataset was obtained from **MLB**, originally curated from the **PitchFx** tracking system. PitchFx records pitch-by-pitch information including pitch velocity, pitch type, release point, and game context. Our dataset aggregates these pitch-level records into player-season summaries.

Career Longevity

- Compare long vs. short MLB career pitchers using **univariate t-tests**
- Fit a **multiple linear regression** model to quantify which early-career variables predict total career length, using **backward stepwise selection** to remove redundant 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 recent pitch-type frequencies and workload.
- Use **Poisson regression** to model discrete performance outcomes (K, BB, HR, outs) and assess how counts of fastballs, breaking balls, and offspeed 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 injured
- **Core Performance:** K/9, BB/9, HR/9
- **Workload:** Velocities of fastball, breaking, offspeed pitches and average pitches per game
- **Style of Pitch:** Proportions of fastball, breaking, and offspeed pitches

To avoid survivorship bias and ensure fair comparisons between players at similar career stages, we restrict our analysis to pitchers who:

- Debuted between **2010–2015**, giving them the opportunity to reach 5+ MLB seasons
- Have at least **two seasons** of pitch-level and performance data

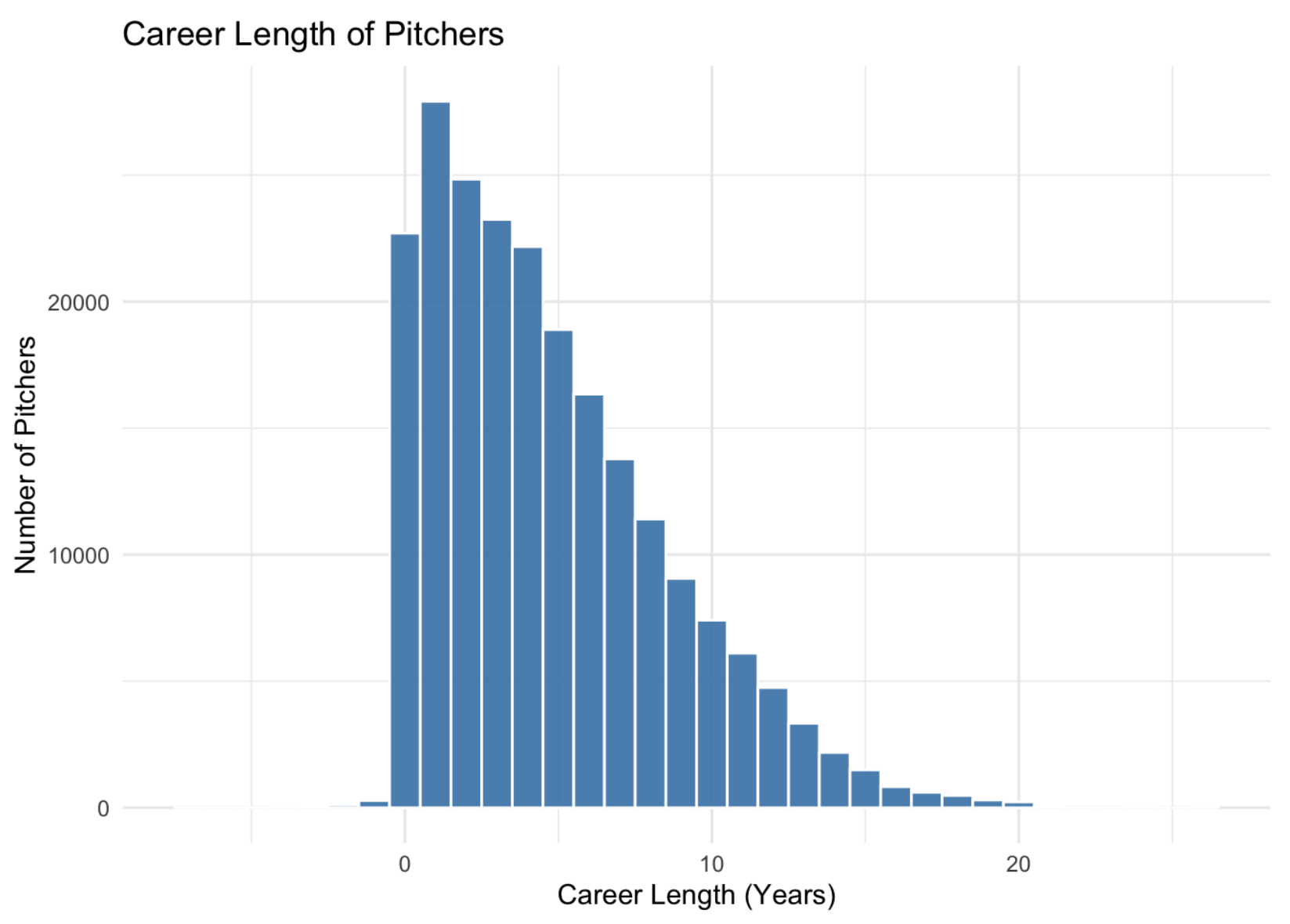

Figure 1. Distribution of total MLB career length (in years) for pitchers

Observations

- The highest concentration of pitchers falls in the **1–3 year** range. This steep early peak reflects how quickly players cycle in and out of the league.
- After approximately **3–4 years**, the number of pitchers who survive beyond this early-career window declines consistently, suggesting that remaining competitive past these thresholds acts as a natural filter for skill, command, health, and durability.
- Only a small fraction of pitchers achieve **10+ year careers**.

Interpretations

This distribution highlights why 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 helps teams distinguish between pitchers who are likely to remain short-term contributors and those who may develop into long-term assets. For organizations, using these insights can improve decisions related to player development, investment, roster planning, and injury management.

1. Univariate Analysis (T-tests)

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

H_0 : All early-career indicators are the same for short and long career pitchers

H_a : At least one indicator differs between groups

Variable	p-value	Mean (Short)	Mean (Long)
Height	0.586	74.28	74.48
Weight	0.734	213.16	214.40
Total Days Injured	0.563	4.19	5.95
K/9	0.287	7.00	7.57
BB/9	**0.017**	4.79	3.62
HR/9	0.095	1.38	0.99
Avg Fastball Velocity	**0.001**	90.98	92.43
Avg Breaking Velocity	0.134	80.49	81.41
Avg Offspeed Velocity	**0.017**	83.00	84.33
Avg Pitches per Game	0.054	33.76	42.90
Prop Fastball	0.255	0.57	0.59
Prop Breaking	0.814	0.21	0.20
Prop Offspeed	0.633	0.09	0.08

Table 1. T-test Results

The only statistically significant predictors of long career length ($p < 0.05$) were:

- **BB/9**: Pitchers who walk fewer hitters last longer. Good command is a key indicator of sustained MLB success.
- **Avg Fastball Velocity**: Higher fastball velocity strongly predicts career longevity. A harder fastball keeps pitchers competitive.
- **Avg Offspeed Velocity**: Higher offspeed velocity also correlates with longer careers, suggesting better mechanics and pitch quality across the arsenal.

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

2. Multiple Linear Regression

A backward stepwise selection procedure identifies three key predictors of career length:

$$\text{Career Length} \sim \text{Total Outs} + \text{BB/9} + \text{Avg Fastball Overall}$$

Variable	Estimate	Std. Error	p-value
Intercept	-44.20	9.10	3.79×10^{-6}
Total Outs	0.00684	0.00139	2.82×10^{-6}
BB/9	-0.244	0.116	0.0367
Avg Fastball Velocity	0.533	0.099	3.95×10^{-7}

Model Fit: $R^2 = 0.383$, Adjusted $R^2 = 0.367$, Residual SE = 2.704, F-statistic $p < 10^{-11}$

VIF Values: Total Outs = 1.15, BB/9 = 1.15, Avg Fastball Velocity = 1.00 (No multicollinearity issues.)

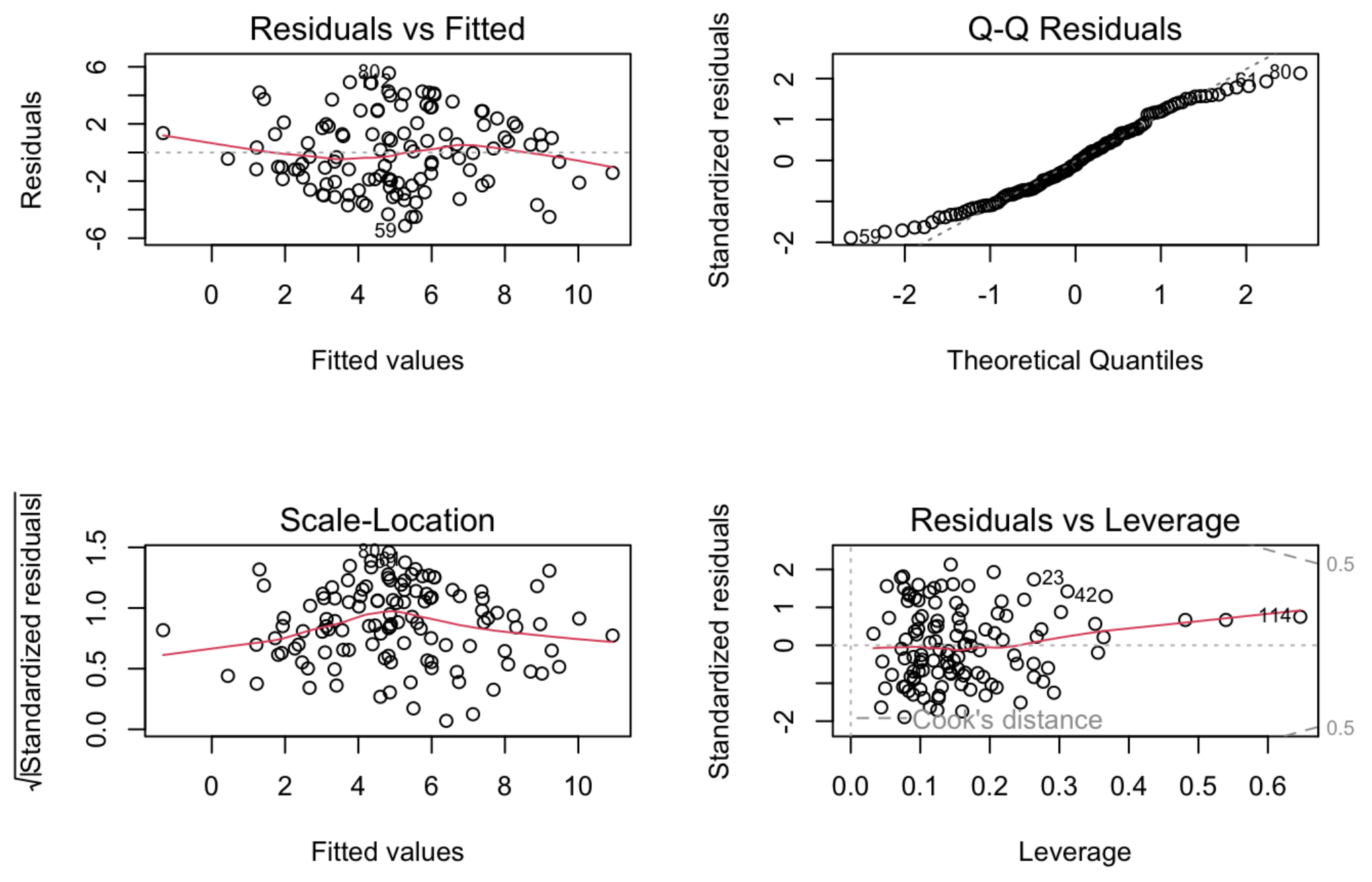

Figure 2. Residuals are roughly normal, evenly distributed, and show no concerning leverage outliers

- **Avg Fastball Velocity** is the strongest positive predictor of career length. Faster pitchers tend to last significantly longer in MLB.
- **BB/9** is negatively associated with career length. Poor command early in a career is a red flag for longevity.
- **Total Outs** reflects a pitcher's workload and skill; early-career durability and success translate into longer careers.

Taken together, the regression suggests that **command + velocity + demonstrated workload** are the most important early indicators of whether a pitcher will have a long MLB career.

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

Background:

To evaluate whether pitch selection contributes to injury, we examine pitchers in their **debut season**. Using only new players helps avoid confounding effects from older, established pitchers who naturally have higher injury risk. For each pitcher, we compute:

- Total fastballs, breaking balls, and offspeed pitches thrown
- Pitch counts from their **last two games** (3-game cumulative count), capturing cumulative workload
- Number of rest days since their previous game

We then model whether the player is placed on the injury list within the next 14 days.

Logistic Regression Model on Injury

$$\text{Injury} \sim \text{3 Game Fastball Count} + \text{3 Game Breaking Count} + \text{3 Game Offspeed Count} + \text{Rest Days}$$

Variable	Estimate	Std. Error	p-value
Intercept	-5.765	0.535	$< 2 \times 10^{-16}$
3 Game Fastball Count	0.0297	0.00979	**0.0024**
3 Game Breaking Count	-0.0210	0.0237	0.377
3 Game Offspeed Count	-0.0315	0.0332	0.343
Rest Days	0.0188	0.0184	0.307

Model Fit: Residual deviance = 254.30, AIC = 264.3.

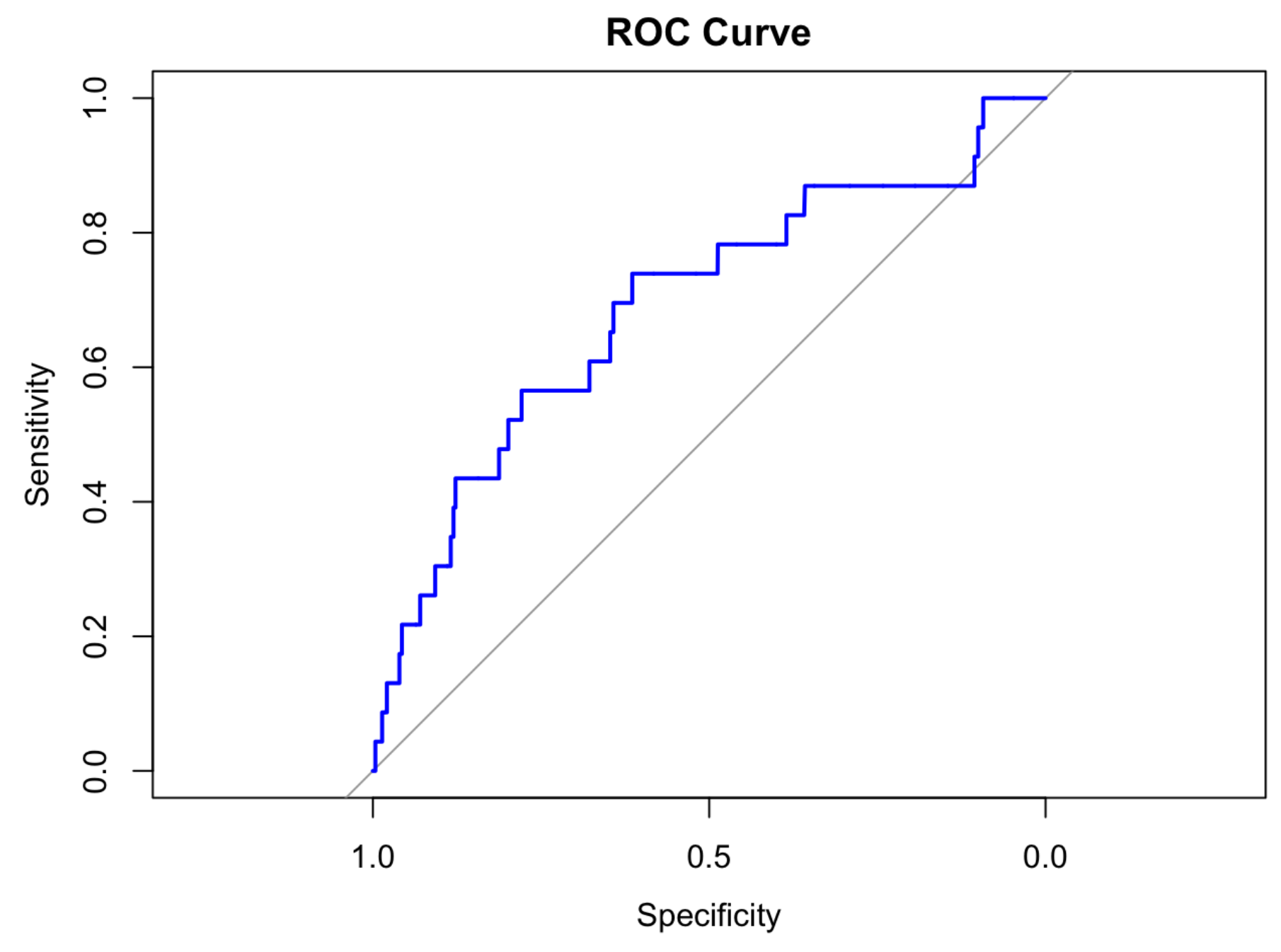

Figure 3. Area Under Curve: 0.6895

The logistic model achieves an **AUC of 0.6895**, indicating moderate discriminatory ability. The model performs noticeably better than random guessing (AUC = 0.5), but does not provide strong predictive accuracy, which is expected given the rarity and noise of short-term injuries

Observations:

- **Fastball frequency is the only significant predictor of short-term injury.** Each additional fastball thrown over the past three games increases the likelihood of injury by approximately **3%**.
- Breaking and offspeed pitches show **no significant association** with injury risk.
- Rest days are not significant; pitchers may recover quickly regardless of spacing between outings.

Interpretation:

- **Fastballs** impose the **greatest biomechanical load** on the arm—higher force production and arm speed—which explains their strong statistical relationship with injury. Even small increases accumulate quickly for players with high fastball usage.
- **Breaking and offspeed pitches** may not register in early-career injuries as they generally generate **lower peak stress** or are **thrown less frequently** by new players.

Conclusion: A heavy fastball workload is a clear early warning signal for short-term injury among new MLB pitchers.

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

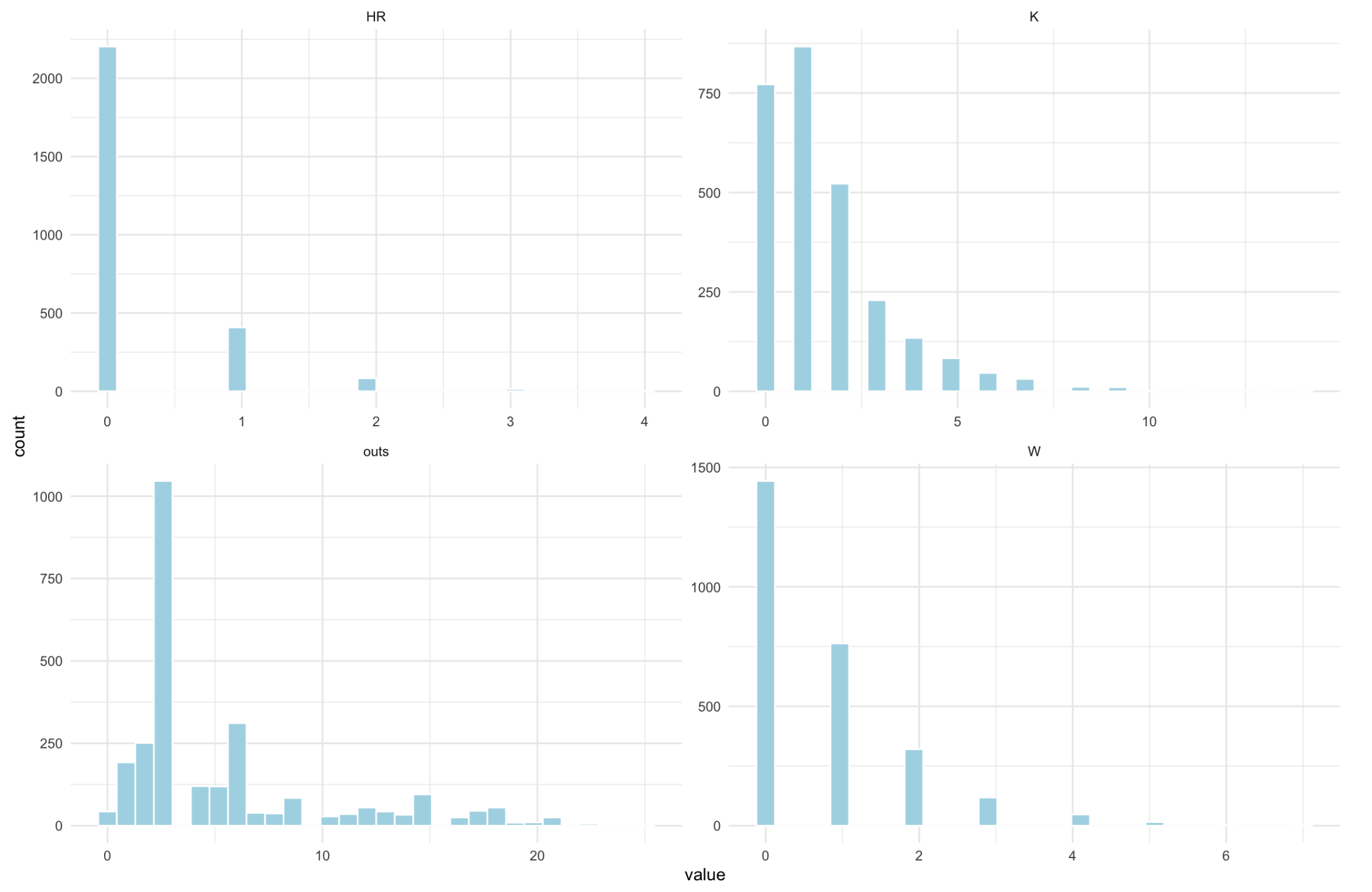

Figure 4. Distribution of performance metrics

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**

Poisson Regression for Performance

$$\log\left(E[\text{outcome}]\right) = \beta_0 + \beta_1 \cdot \text{fastball} + \beta_2 \cdot \text{breaking} + \beta_3 \cdot \text{offspeed}.$$

Each coefficient represents the change in the log expected count of the outcome for each additional pitch of that type.

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	5.61×10^{-16}

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	3.23×10^{-16}
Offspeed Count	0.0137	0.003470	7.78×10^{-5}

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	1.11×10^{-15}
Breaking Count	0.0325	0.004235	1.59×10^{-14}
Offspeed Count	0.0386	0.005669	9.26×10^{-12}

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	$< 2 \times 10^{-16}$

Model Results

All four Poisson models (K, W, HR, outs) significantly outperform the intercept-only model based on deviance and AIC. All coefficients are positive and statistically significant ($p < 0.001$), but their magnitudes vary by outcome.

- **Strikeouts (K): Breaking balls have the strongest effect.** The breaking coefficient is the largest among the three ($\beta = 0.0284$), indicating that breaking pitches are the most effective at generating strikeouts.
- **Walks (W): Fastballs increase walk probability the most.** Fastballs have the highest coefficient ($\beta = 0.0251$), suggesting that fastballs are harder to locate consistently, especially for new players.
- **Home Runs (HR): Offspeed pitches are most strongly associated with HRs.** Offspeed has the largest coefficient ($\beta = 0.0386$), aligned with baseball intuition: poorly executed offspeed pitches get hit hard.
- **Outs: All three pitch types have similar effects.** Coefficients for outs are nearly identical across pitch types, indicating that no single pitch type dominates in generating outs.

Conclusion: Different pitch types specialize in different outcomes. Breaking balls are best for strikeouts, fastballs carry higher walk risk, offspeed pitches pose home-run danger, and outs appear largely pitch-agnostic.

Outcome	Most Associated Pitch Type	Reason / Interpretation
Strike (K)	Breaking Ball	Highest coefficient; hardest pitch type to hit
Walk (W)	Fastball	Harder to control at high velocity
Home Run (HR)	Offspeed	Easiest to hit if anticipated or poorly located
Out	All Similar	Coefficients nearly equal across pitch types

Table 2. Summary of Poisson Coefficient Interpretations for Each Outcome

Conclusion

Key Findings

- **Longevity (RQ1):** Longer careers are associated with lower BB/9, higher fastball velocity, and early-career workload.
- **Injury Risk (RQ2a):** Only recent fastball volume predicts injury; breaking/offspeed counts are not significant.
- **Performance (RQ2b):** Breaking balls → most strikeouts; fastballs → most walks; offspeed → most HRs; outs ≈ similar across pitch types.

Limitations

- **Seniority inconsistencies:** Some players show negative or irregular seniority values due to pre-MLB practice or offseason games, which may distort early-career statistics.
- **Censoring of career length:** Data ends in 2020, so pitchers labeled as "short-career" may still be active or returning from injury, leading to underestimation of true longevity.
- **Injury-list ambiguity:** Injury List (IL) placement is not always strictly medical; teams may use it for strategic rest, roster flexibility, or minor, undocumented issues, which introduces noise into the injury outcome.
- **Unobserved confounding:** We lack biomechanics, training load, medical history, and conditioning data, all of which influence both injury risk and long-term development.

Recommendations

Injury Prevention

Teams should closely monitor rolling fastball workload, as sustained stretches of high fastball usage meaningfully increase short-term injury risk. Real-time pitch-mix tracking can help identify emerging fatigue before it becomes harmful. Recovery plans should be tailored to recent pitch volume rather than relying solely on rest-day counts, which often fail to reflect true physiological stress.

Player Development

Early-career development should emphasize sharpening command, since lower walk rates consistently predict longer MLB careers. Improving breaking-ball quality offers the greatest upside for generating strikeouts, while refining offspeed control can directly reduce susceptibility to home runs. Together, these adjustments help pitchers balance effectiveness with long-term durability.

Scouting and Evaluation

Scouts should prioritize identifying prospects with strong fastball quality as this pitch is the foundation of pitching success. A well-executed fastball tends to drive better performance outcomes across strikeouts, weak contact, and overall run prevention, which in turn is strongly associated with longer career longevity. Because pitchers with effective fastballs are more likely to remain competitive at higher levels, scouts should focus on evaluating fastball strength, command, and sustainability.

https://statistics.ucla.edu/ Fall 2025 STATS 140XP vlew@stat.ucla.edu