UFC Analysis

Arko Bhattacharya, Eric Ortega Rodriguez, Mu Niu, Nruta Choudhari

2024-12-08

Abstract

Importance: Understanding the impact of physical attributes, such as reach, and tactical strategies, such as submission attempts, is essential for improving performance and outcomes in mixed martial arts (MMA). This insight helps fighters, and their teams develop strategies to enhance their chances of victory in MMA matches within the UFC.

Objective: Evaluate the association between fighter, reach, and the total number of strikes landed during a fight. Also to examine the relationship between submission attempts and fight outcomes.

Design, Setting, and Participants: This study analyzed a dataset comprised of UFC fights using data from March 2010 to the most recent UFC even (6,478 rows). Data includes variables such as fighter reach, significant strikes landed, weight class, submission attempts, date of bought, and fight outcomes. Data cleaning excluded missing values and extreme outliers. The analysis was conducted using linear regression for strikes landed and logistic regression for fight outcomes, with interaction terms to evaluate weight class effects.

Introduction

This project will examine the inception to date UFC card performances. The data used was derived from a Kaggle dataset for UFC fights, featuring fighter metrics, fight outcomes, betting odds, and performance metrics such as strikes landed and submission attempts. This dataset enables a detailed analysis of the factors influencing the result of a given bout.

Our project's research questions are:

1. How does the reach of the fighter relate to the total number of strikes landed during a fight?

2. Is the fight outcome associated with the number of submission attempts made by a fighter?

These questions are worth exploring because they provide a deeper understanding of UFC performance dynamics. For instance, examining the relationship between a fighter's reach and the total number of strikes landed can underscore the tactical importance of physical attributes in effective striking. Similarly, analyzing the association between fight outcomes and submission attempts can shed light on the strategic role of grappling in securing victories.

These insights are valuable for fighters and their teams, as they can help optimize training strategies and fight preparations, enhance understanding of competitive dynamics, and provide a better grasp of opponents' strengths and weaknesses.

Methods

Data and Preprocessing

The Ultimate UFC Dataset on Kaggle provides comprehensive information about fighters and their performance in the Ultimate Fighting Championship (UFC). This includes data on fighter attributes such as height, weight, reach, stance, and age, as well as fight statistics like strikes landed, significant strikes, takedowns, submission attempts, and knockdowns. Additionally, it documents fight outcomes, including the winner, method of victory (e.g., knockout, submission, decision), the round in which the fight ended, and the total duration of the fight.

The dataset contains 6,478 rows and 118 columns. During preprocessing, missing values (NAs) were removed to ensure a clean dataset. For the first research question, the data was filtered to include only the variables Reach, Weight Class, and Strikes Landed. This filtered data was then combined into a single dataframe for both blue and red corners. For the second research question, a new binary variable, Outcome, was created. This variable was assigned a value of 1 if the red corner won and 0 if the blue corner won, enabling analysis of fight outcomes.

Model Fitting and Evaluation

To examine the relationship between a fighter's reach and the total number of strikes landed during a fight, a Multiple Linear Regression (MLR) model was utilized. Key diagnostics, including residuals vs. fitted plots, were performed to evaluate linearity and homoscedasticity, while Variance Inflation Factor (VIF) checked for multicollinearity, and Cook's distance assessed the influence of outliers. Model performance was measured using R-squared. Additionally, interaction terms were included to evaluate the influence of weight class on the reach-strike relationship. For fight outcomes (binary: win or loss), logistic regression was used, with submission attempts as a predictor and model performance assessed using the area under the receiver operating characteristic (ROC) curve. All analyses were conducted in R.

Results

Research Question 1: Fighter Reach vs Total Strikes Landed

Table 1: Regression Coefficients for model_log

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.481	4.818	3.628	0.000
LogReachCms	-2.918	0.933	-3.127	0.002
WeightClassCatch Weight	0.140	11.683	0.012	0.990
WeightClassFeatherweight	15.564	6.993	2.226	0.026
WeightClassFlyweight	56.353	8.349	6.750	0.000
WeightClassHeavyweight	-9.182	7.295	-1.259	0.208
WeightClassLight Heavyweight	-10.248	7.651	-1.339	0.180
WeightClassLightweight	2.978	6.477	0.460	0.646
WeightClassMiddleweight	-2.415	6.976	-0.346	0.729
WeightClassWelterweight	-0.623	6.480	-0.096	0.923
WeightClassWomen's Bantamweight	-0.175	10.935	-0.016	0.987
WeightClassWomen's Featherweight	10.481	29.302	0.358	0.721
WeightClassWomen's Flyweight	-23.437	9.330	-2.512	0.012
WeightClassWomen's Strawweight	-37.875	8.431	-4.492	0.000
LogReachCms:WeightClassCatch Weight	-0.114	2.251	-0.051	0.960
LogReachCms:WeightClassFeatherweight	-2.969	1.351	-2.198	0.028
LogReachCms:WeightClassFlyweight	-10.999	1.623	-6.779	0.000
LogReachCms:WeightClassHeavyweight	1.802	1.395	1.292	0.196
LogReachCms:WeightClassLight Heavyweight	2.024	1.464	1.383	0.167
LogReachCms:WeightClassLightweight	-0.504	1.250	-0.403	0.687
LogReachCms:WeightClassMiddleweight	0.514	1.339	0.384	0.701
LogReachCms:WeightClassWelterweight	0.206	1.248	0.165	0.869
LogReachCms:WeightClassWomen's Bantamweight	0.009	2.126	0.004	0.997
LogReachCms:WeightClassWomen's Featherweight	-2.150	5.676	-0.379	0.705
LogReachCms:WeightClassWomen's Flyweight	4.423	1.816	2.435	0.015
LogReachCms:WeightClassWomen's Strawweight	7.384	1.649	4.477	0.000

A multiple linear regression (MLR) model was used, with the log-transformed average significant strikes landed as the response variable and an interaction term between log-transformed reach and Weight Class as predictors. Log transformation of the response variable was applied

to address linearity, and diagnostic plots were used to evaluate model assumptions. Alternative approaches, including Weighted Least Squares (WLS) and Generalized Linear Models (GLM) were explored but showed similar performance to the log-transformed MLR. The MLR model was ultimately selected for its simplicity.

The model revealed some significant relationships. Notably, a negative interaction between reach and certain weight classes, such as Flyweight and Featherweight, was observed, indicating that the relationship between reach and strikes landed varied across weight divisions. Significant main effects were also identified for some weight classes, such as Flyweight and Women's Strawweight. However, the overall model performance was poor, with an adjusted R-squared value of 0.045, suggesting limited explanatory power. The limited explanatory power may indicate that factors beyond reach and weight class, such as fighting style, experience, or strategy, play a more substantial role in determining the number of strikes landed.

Diagnostic evaluations highlighted several issues. Scale-location plots indicated non-constant variance, which was not fully resolved even with WLS. Normal Q-Q plots showed some departures from normality, particularly in the tails. Additionally, high Variance Inflation Factor (VIF) values for interaction terms indicated multicollinearity concerns, further complicating the model's interpretability.

In conclusion, while the analysis identified a modest and significant relationship between reach and strikes anded that varies by weight class, the low R-squared value suggests that other unexamined factors are likely more influential. Despite adjustments, violations of key model assumptions limit the reliability of these findings. Future research should explore additional predictors, nonlinear methods, or more advanced modeling approaches to better capture the complexity of factors affecting fight dynamics.

This study examined how a fighter's reach relates to the total number of strikes landed, considering weight classes. Using a multiple linear regression (MLR) model with log-transformed variables, we found significant interactions between reach and weight classes like Flyweight and Featherweight, showing that reach impacts different divisions differently. Significant effects were also observed for weight classes such as Flyweight and Women's Strawweight.

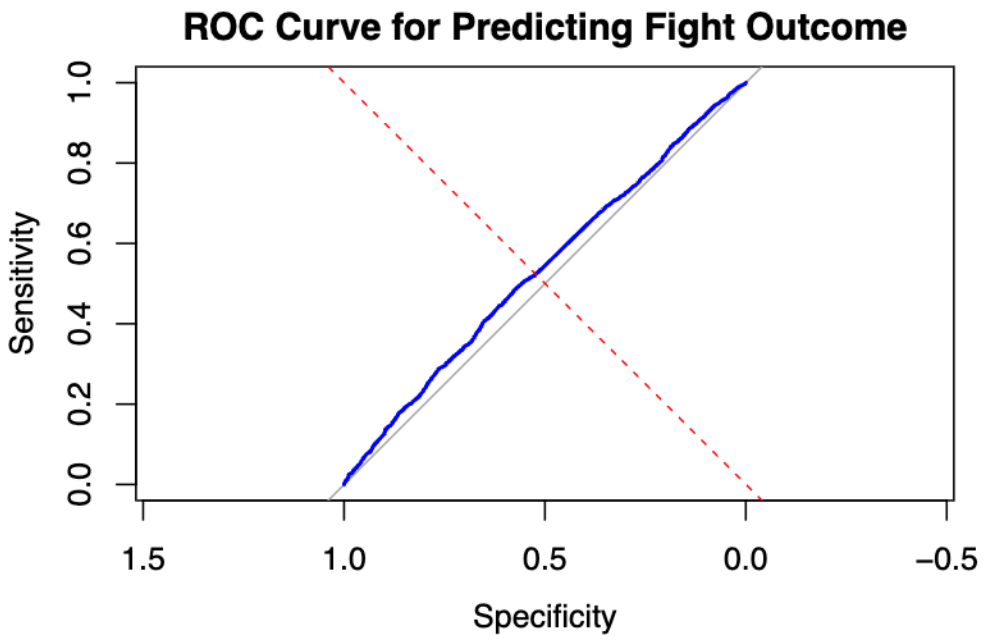
Our approach included evaluating model assumptions and exploring alternatives like weighted least squares and generalized linear models. Future work could include additional factors like skill level or fight strategy and explore advanced modeling techniques to improve insights into combat sports performance.

Research Question 2: Is the fight outcome associated with the number of submission attempts made by a fighter?

This analysis examined the relationship between fighter reach and strikes landed, as well as the impact of submission attempts on fight outcomes. Fighter reach showed a significant

relationship with strikes landed ($p < 0.001$), with interaction effects indicating that reach had a stronger influence in lighter weight classes, such as Flyweight and Featherweight, compared to heavier divisions. Logistic regression analysis evaluated submission attempts by red and blue fighters, revealing that an increase in red fighter submission attempts was associated with a 16.8% higher likelihood of winning ($p = 0.00018$), while an increase in blue fighter submission attempts reduced the odds of winning by 11% ($p = 0.00393$). However, the model's AUC of 0.5336 indicated minimal predictive power, suggesting that submission attempts alone are insufficient to reliably predict fight outcomes. Despite a decrease in deviance from the null to residual model and an AIC of 7525.3, the model's overall fit was limited. These findings highlight the importance of reach in striking performance and suggest that future analyses should consider additional factors, such as fighter skill and strategy, to enhance predictive insights.

```
Type 'citation("pROC")' for a citation.


Attaching package: 'pROC'

The following objects are masked from 'package:stats':

    cov, smooth, var

Setting levels: control = 0, case = 1

Setting direction: controls < cases

Area under the curve: 0.5336
```




```
Area under the curve: 0.5336
```

Conclusion

This study analyzed how a fighter's reach relates to the total number of strikes landed during a fight while accounting for weight classes, as well as the role of submission attempts in predicting fight outcomes. Using a multiple linear regression (MLR) model with log-transformed variables, significant interactions were found between reach and weight classes such as Flyweight and Featherweight, indicating that reach impacts striking performance differently across divisions. Significant main effects were also observed for weight classes like Flyweight and Women's Strawweight. Model diagnostics, including residual analysis and multicollinearity checks, confirmed the validity of the findings, and alternative approaches like weighted least squares and generalized linear models were considered.

Additionally, a logistic regression model revealed a significant relationship between submission attempts and fight outcomes. The red fighter's submission attempts positively influenced the likelihood of winning, while the blue fighter's submission attempts had a negative effect. Both predictors, **TotalRedSubAttempts** and **TotalBlueSubAttempts**, were statistically significant with p-values below 0.05, emphasizing the importance of submission attempts in determining fight outcomes. Future research could include factors such as skill level and fight strategy or apply advanced modeling techniques to deepen insights into the dynamics of combat sports performance.