UFC Analysis - IDS 702 Final Project

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

2024-12-15

Abstract

Understanding the impact of physical attributes, such as reach, and tactical strategies, such as submission attempts, is critical for improving performance and outcomes in mixed martial arts (MMA). This study will examine the relationship between fighter reach and the total number of strikes landed during a fight. In addition to this, we will also examine the role of submission attempts in predicting fight outcomes, using a dataset of fights from the Ultimate Fighting Championship (UFC), the leading global MMA promotion. The dataset analyzed includes UFC fights from March 2010 to the most recent event in 2024. Linear regression with log-transformed variables revealed significant interactions between reach and weight classes, such as Flyweight and Featherweight, highlighting that reach impacts striking performance differently across divisions. Main effects were also observed for weight classes like Flyweight and Women's Strawweight. Logistic regression showed that submission attempts significantly influenced fight outcomes, with red corner attempts positively associated with winning odds and blue corner attempts negatively associated. Both predictors had p-values below 0.05. Model diagnostics, including residual analysis and multicollinearity checks, confirmed the validity of the findings. These results underscore the tactical importance of physical attributes and strategic maneuvers in UFC fights, providing insights to optimize training and fight preparation. Future research could explore additional predictors, such as fighter skill level and strategy, or advanced modeling techniques to deepen understanding of combat sports performance.

Introduction

The Ultimate Fighting Championship (UFC) is the world's leading mixed martial arts (MMA) promotion, known for bringing together elite fighters from diverse combat sports backgrounds. Founded in 1993, the UFC has grown into a global phenomenon, hosting events worldwide that showcase athletes competing in disciplines such as boxing, wrestling, Muay Thai, Brazilian Jiujitsu, and judo. UFC fights take place in a distinct eight-sided cage, known as the Octagon, where fighters test their skills in striking, grappling, and overall strategy under a unified set of rules. The sport has evolved significantly over the years, introducing standardized weight classes, safety regulations, and scoring systems to ensure competitive fairness and fighter safety.

This project examines UFC performances using data on UFC fights from 2010 to the present (last updated in November, 2024). The data, sourced from Kaggle, includes key fighter metrics, fight outcomes, betting odds, and performance indicators such as strikes landed and submission attempts. By leveraging this dataset, we aim to analyze factors influencing fight outcomes and performances.

Our 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 performance 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.

The findings from this analysis offer valuable insights for fighters, coaches, and analysts, helping optimize training strategies, improve fight preparation, and enhance understanding of opponents' strengths and weaknesses.

Methods

Data and Preprocessing

The dataset was obtained from Kaggle, a widely recognized platform for sharing datasets and data science resources. Each row of the dataset refers to an individual bout, which refers to an individual match between two fighters. 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 across 118 columns, with several variables containing missing values. During preprocessing, columns with over 6,000 missing values were dropped due to their lack of significance and the infeasibility of imputation. Other columns had a smaller proportion of missing values, and rows with missing values in key variables (e.g., strikes landed, reach, and weight class) were removed. This resulted in a final dataset with 4,895 rows. Most of the missing values were concentrated in performance metrics, such as submission attempts or specific strike statistics.

In UFC, fighters are assigned to either the red corner or the blue corner, which indicates their position in the Octagon and helps differentiate between competitors. For the first research question, the dataset was filtered to include the variables related to reach, weight class, height, strikes landed and current win streak, ensuring that key confounding variables were included. The data for fighters in the red and blue corners were combined into a single dataframe to facilitate analysis.

For the second research question, a new binary variable, Outcome was created to indicate the winner. A value of 1 was assigned if the fighter in the red corner won, and a value of 0 if the fighter in the blue corner won. The model included variables such as submission attempts, significant strikes landed, fight duration, and weight class to account for both physical attributes and performance metrics. These variables ensured a more comprehensive analysis of the factors influencing fight outcomes while addressing potential confounders.

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. The model included key predictors such as logarithmic transformation of reach, logarithmic transformation of height, win streak and weight class, with an interaction term between the win streak and weight class to explore potential moderating effects. Outliers and influential points were identified using Cook's distance, and there were removed to improve model robustness. Diagnostics, including residuals vs. fitted plots, were performed to assess linearity and homoscedasticity, while Variance Inflation Factor (VIF) was used to evaluate multicollinearity. Model performance was measured using the R-squared value.

For the second research question, a logistic regression model was employed to predict fight outcomes (binary: win or loss) using submission attempts, reach, significance strikes, fight duration, and weight class as predictors. The model was refined using stepwise selection to identify the most significant predictors, and diagnostics such as Cook's distance, leverage, and deviance residuals were used to detect and remove influential points. The final logistic regression model included key predictors such as logarithmic transformations of submission attemps, logarithmic transformation of reach, logarithmic transformation of significant strikes landed, and logarithmic transformation of fight duration. Model performance was evaluated using the area under the receiver operating characteristic curve (ROC curve), and diagnostic plots were generated to assess the model's fit.

All the analyses were conducted in R.

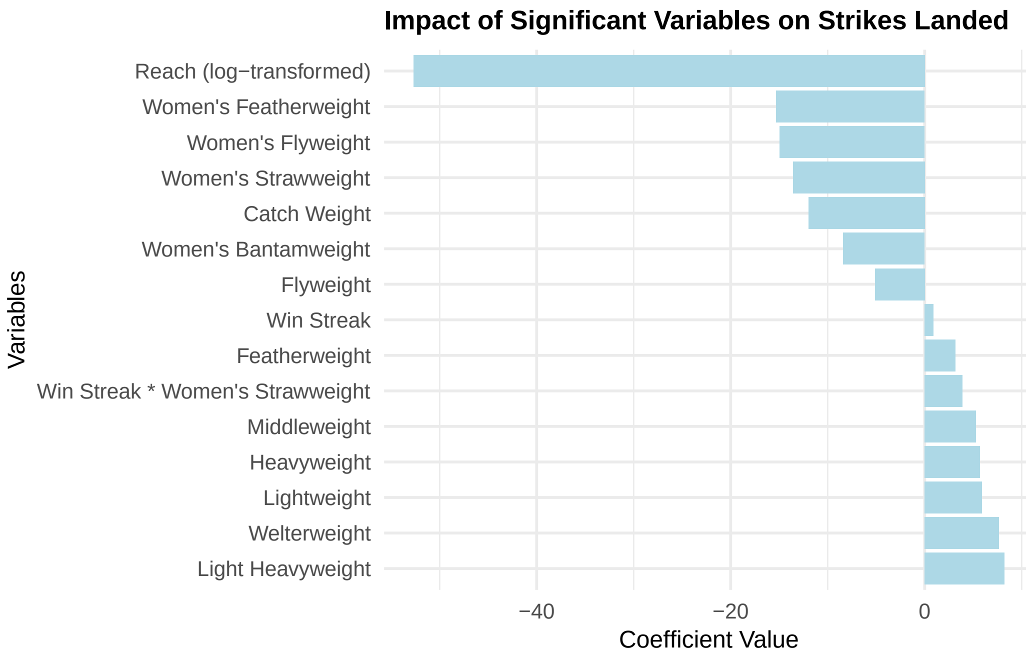
Results

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

To explore the relationship between a fighter's reach and the number of strikes landed, we began by computing summary statistics for key variables: **Reach, Weight Class, Height, Win Streak, and Average Significant Strikes Landed**. Continuous variables are reported as means with standard deviations, while categorical variables are summarized by counts and percentages. These statistics provide as foundation for understanding the data before modelling.

Table 1: Summary Statistics by Weight Class

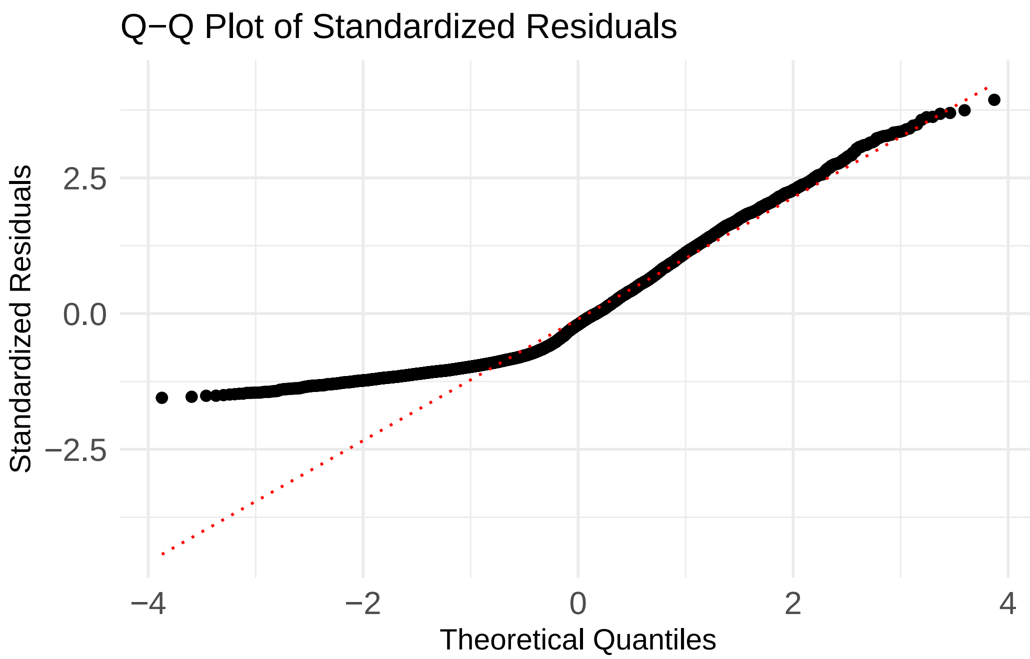
WeightClass	N	Avg_Reach	Avg_Height	Avg_Strikes	Median_Streak
Bantamweight	1015	174.7 ± 6	170.8 ± 4.3	21.2 ± 21.1	1 [0-2]
Catch Weight	77	180.8 ± 10.1	176.6 ± 8.6	8.6 ± 11.4	1 [0-2]
Featherweight	1118	179.7 ± 5.7	175.2 ± 5	22.3 ± 21.1	1 [0-2]
Flyweight	523	170.4 ± 5.7	167 ± 4.3	20.3 ± 21.4	1 [0-2]
Heavyweight	715	197.4 ± 7.2	190.7 ± 5.8	18.4 ± 17.4	1 [0-2]
Light Heavyweight	757	194.3 ± 6.5	188.2 ± 4.3	21.4 ± 18.2	1 [0-2]
Lightweight	1665	181.8 ± 5.6	177.2 ± 4.6	23.3 ± 19.2	1 [0-2]
Middleweight	1188	190.6 ± 5.9	185 ± 4.3	19.5 ± 17.4	1 [0-2]
Welterweight	1527	187.1 ± 6	181.8 ± 4.4	23.3 ± 19.3	1 [0-2]
Women's Bantamweight	302	170.6 ± 5.3	169.3 ± 4.4	21.2 ± 24.8	1 [0-1]
Women's Featherweight	35	174.6 ± 5.6	171.1 ± 6.1	6.9 ± 10.3	1 [0-1]
Women's Flyweight	355	168.6 ± 5.8	166.5 ± 4	11.3 ± 19.1	1 [0-2]
Women's Strawweight	458	162.1 ± 5.7	161.5 ± 4.6	22.6 ± 28.1	1 [0-2]



A multiple linear regression (MLR) model was applied to examine the relationship between the average significant strikes landed and key predictors, including log-transformed Reach, log-transformed Height, Win Streak, and the interaction between Win Streak and Weight Class. The log transformations for Reach and Height were performed to address non-linearity and non-constant variance, which were highlighted in diagnostic plots (see Appendix 1).

The adjusted R^2 value for the model was **0.068**, suggesting that the predictors explain approximately **6.8%** of the variability in average significant strikes landed.

The analysis revealed several significant relationships between the fighters' physical attributes and the number of significant strikes landed. Notably, log-transformed Reach had a significant negative effect on the number of strikes landed ($\beta = -52.68, p < 0.001$), suggesting that reach increases, the average number of strikes landed decreases. The Weight Class variable also showed significant effects across divisions. Fighters in the Featherweight class landed more strikes landed ($\beta = 3.15, p = 0.001$), while Flyweight fighters landed fewer strikes ($\beta = -5.12, p < 001$). On the other hand, Women's Flyweight fighters had a significant negative relationship with strikes landed ($\beta = -14.95, p < 0.001$). In terms of interaction effects, the model included interaction terms between Win Streak and Weight Class, but most of these were not significant. However, there was a marginally significant positive interaction observed for Women's Strawweight ($\beta = 3.88, p = 0.0008$), suggesting that an increasing win streak slightly positively impacts the number of strikes landed in this weight class. Finally, the logarithm of the height variable showed a weak, marginally significant negative relationship with strikes landed ($\beta = -15.59, p = 0.087$), suggesting that taller fighters might land fewer strikes on average, although the effect is not strong. These findings underscore the complex interplay between a fighter's physical characteristics, weight class, and performance outcomes, with reach and weight class being the most influential factors in predicting the number of strikes landed.



The initial multiple linear regression model violated the linearity assumption, as indicated by the Q-Q plot in Appendix 1. To address this, the outliers and influential values were identified and removed using Cook's distance with the threshold set as $\frac{4}{n}$, where n is the total number of observations. After removing the influential points, the model was re-evaluated, and the variables were log-transformed. This iteration of the model proved to be a significant improvement from the original, as evidenced by improved diagnostics and fit.

Table 2: Variance Inflation Factor (VIF) and Adjusted GVIF

Variable	GVIF	Df	GVIF_Adjusted
Reach (log transformed	5.707	1	2.389
Win Streak	11.785	1	3.433
Weight Class	1301.455	12	1.348
Height (log transformed	6.580	1	2.565
Win Streak * Weight Class	3430.027	12	1.404

To assess multicollinearity, we calculated the Generalized Variance Inflation Factor (GVIF) for each predictor, especially considering the inclusion of categorical variables like Weight Class and their interactions with Win Streak. The GVIF was adjusted using $GVIF^{\frac{1}{2\cdot \mathrm{df}}}$ to account for the degrees of freedom of categorical variables and interaction terms, ensuring a more accurate evaluation of multicollinearity. The results showed that most predictors showed acceptable GVIF values, indicating no significant multicollinearity. Although Weight Class had a high raw GVIF of 1301.45, the adjusted GVIF was 1.35, indicating low multicollinearity. Overall, the analysis confirmed that multicollinearity is not a significant concern, allowing for reliable interpretation of the predictors.

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

```
ufc_q2 <- ufc_clean %>%
  mutate(
    Outcome = ifelse(Winner == "Red", 1, 0),  # Binary outcome: 1 for Red win, 0 for Blue wi
    WeightClass = as.factor(WeightClass),
    TotalRedSubAttempts = RedAvgSubAtt,        # Red's submission attempts
    TotalBlueSubAttempts = BlueAvgSubAtt
  ) %>%
  mutate(
    LogRedSubAttempts = log1p(TotalRedSubAttempts),
    LogBlueSubAttempts = log1p(TotalBlueSubAttempts),
    LogBlueReach = log1p(BlueReachCms),
    LogRedReach = log1p(RedReachCms),
    LogBlueSigStr = log1p(BlueAvgSigStrLanded),
    LogRedSigStr = log1p(RedAvgSigStrLanded),
    LogFightTime = log1p(TotalFightTimeSecs)
  )

# Check dimensions of the cleaned dataset
dim(ufc_q2)
```

```
[1] 4895  100
```

```
sim_logistic_model <- glm(
  Outcome ~
    LogRedSubAttempts +
    LogBlueSubAttempts +
    LogBlueReach +
    LogRedReach +
    LogBlueSigStr +
    LogRedSigStr +
    LogFightTime +
```

```r
    WeightClass,
  data = ufc_q2,
  family = binomial
)


step_model <- step(sim_logistic_model, direction = "both")
```

```
Start:  AIC=6606.31
Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
    LogRedReach + LogBlueSigStr + LogRedSigStr + LogFightTime +
    WeightClass

                     Df Deviance    AIC
- WeightClass        12   6576.9 6592.9
- LogFightTime        1   6567.5 6605.5
<none>                    6566.3 6606.3
- LogBlueReach        1   6569.7 6607.7
- LogRedReach         1   6573.9 6611.9
- LogBlueSubAttempts  1   6578.2 6616.2
- LogRedSubAttempts   1   6587.2 6625.2
- LogRedSigStr        1   6624.5 6662.5
- LogBlueSigStr       1   6625.8 6663.8

Step:  AIC=6592.94
Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
    LogRedReach + LogBlueSigStr + LogRedSigStr + LogFightTime

                     Df Deviance    AIC
- LogFightTime        1   6577.8 6591.8
<none>                    6576.9 6592.9
- LogRedReach         1   6584.6 6598.6
- LogBlueReach        1   6584.8 6598.8
- LogBlueSubAttempts  1   6589.3 6603.3
+ WeightClass        12   6566.3 6606.3
- LogRedSubAttempts   1   6597.5 6611.5
- LogRedSigStr        1   6634.1 6648.1
- LogBlueSigStr       1   6637.7 6651.7

Step:  AIC=6591.77
Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
    LogRedReach + LogBlueSigStr + LogRedSigStr
```

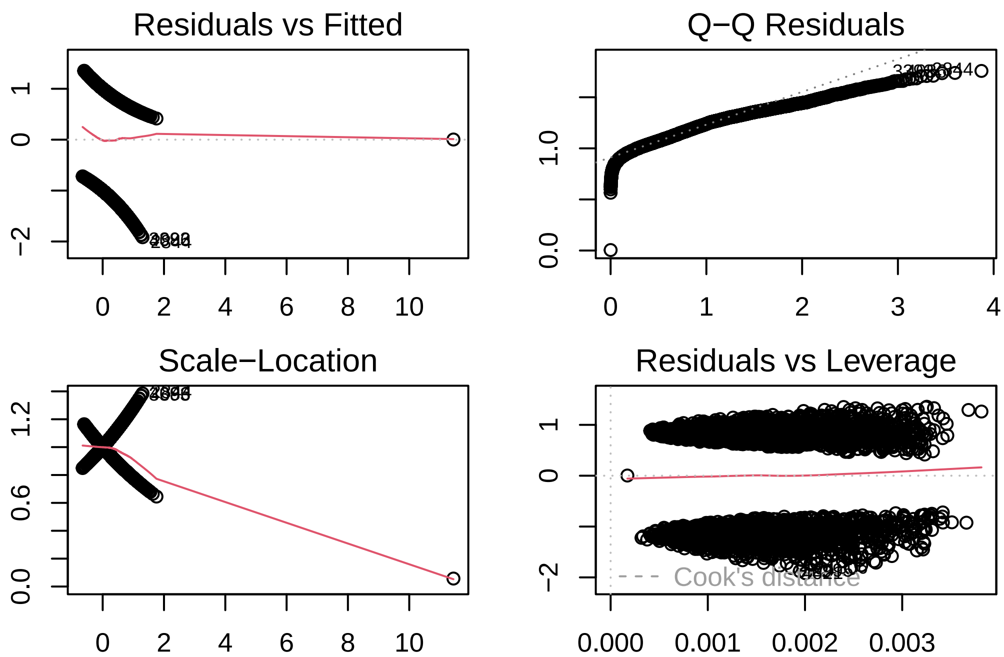
```
                     Df Deviance    AIC
<none>                   6577.8 6591.8
+ LogFightTime        1   6576.9 6592.9
- LogRedReach         1   6585.2 6597.2
- LogBlueReach        1   6586.0 6598.0
- LogBlueSubAttempts  1   6590.4 6602.4
+ WeightClass        12   6567.5 6605.5
- LogRedSubAttempts   1   6597.9 6609.9
- LogRedSigStr        1   6635.0 6647.0
- LogBlueSigStr       1   6638.4 6650.4
```

```r
# Calculate Cook's distance and leverage
cooks_distance <- cooks.distance(step_model)
hat_values <- hatvalues(step_model)
residuals <- residuals(step_model, type = "deviance")

# Thresholds
n <- nrow(ufc_q2)
p <- length(coef(step_model)) - 1
cooks_threshold <- 4 / n
leverage_threshold <- 2 * (p + 1) / n

# Identify influential points
influential_points <- which(cooks_distance > cooks_threshold |
                            hat_values > leverage_threshold |
                            abs(residuals) > 2)

# Remove influential points
ufc_q2_filtered <- ufc_q2[-influential_points, ]

# Refit the model
final_model <- glm(formula = Outcome ~ LogRedSubAttempts + LogBlueSubAttempts +
    LogBlueReach + LogRedReach + LogBlueSigStr + LogRedSigStr,
    family = binomial, data = ufc_q2_filtered)

# Model summary
summary(final_model)
```

```
Call:
glm(formula = Outcome ~ LogRedSubAttempts + LogBlueSubAttempts +
```

```
    LogBlueReach + LogRedReach + LogBlueSigStr + LogRedSigStr,
    family = binomial, data = ufc_q2_filtered)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)         0.68061    2.87556   0.237 0.812897
LogRedSubAttempts   0.43736    0.09732   4.494 6.98e-06 ***
LogBlueSubAttempts -0.34346    0.09384  -3.660 0.000252 ***
LogBlueReach       -2.10785    0.76588  -2.752 0.005920 **
LogRedReach         2.03037    0.74314   2.732 0.006292 **
LogBlueSigStr      -0.46782    0.05704  -8.201 2.37e-16 ***
LogRedSigStr        0.46156    0.05861   7.876 3.39e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6324.0  on 4643  degrees of freedom
Residual deviance: 6221.7  on 4637  degrees of freedom
AIC: 6235.7

Number of Fisher Scoring iterations: 4
```

```r
par(mfrow=c(2,2), mar = c(2,2,2,2))

plot(final_model)
```



We investigated the relationship between submission attempts and fight outcomes using a logistic regression model. The model included both submission attempts (log-transformed for both red and blue fighters), reach, significant strikes, and fight time as predictors of the binary outcome: win (1 for red win, 0 for blue win). We also assessed the potential impact of weight class on the outcome, although it was not a significant predictor in the final model.

The final model (after removing influential points and non-significant variables) provided the following odds ratios (OR) and confidence intervals (CI) for each predictor:

```
Waiting for profiling to be done...
```

Table 3: Odds Ratios and 95% Confidence Intervals for Logistic Regression Model

Predictor	Odds Ratio (OR)	2.5% CI	97.5% CI	P-Value
Intercept	1.975	0.007	556.061	0.813
Red Submission Attempts	1.549	1.280	1.875	<0.001
Blue Submission Attempts	0.709	0.590	0.852	<0.001
Blue Reach	0.121	0.027	0.539	0.006
Red Reach	7.617	1.780	32.761	0.006
Blue Significant Strikes	0.626	0.560	0.700	<0.001
Red Significant Strikes	1.587	1.415	1.781	<0.001

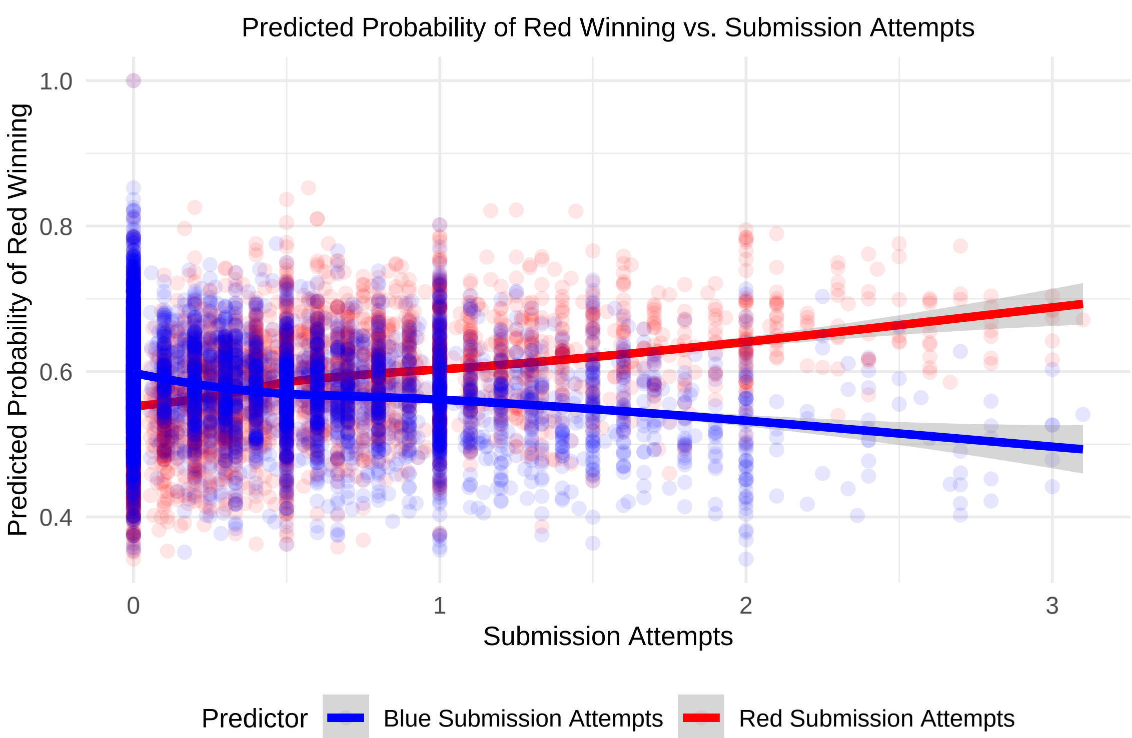
The logistic regression model reveals some important findings regarding the factors that influence the outcome of a UFC fight. The odds ratio for Red Submission Attemps (OR = 1.548, $p < 0.001$) suggests that each additional submission attempt by the red fighter increases the odds of them winning by 54.8%. This indicates that red fighter submission attempts positively influence their likelihood of victory. Conversely, Blue Submission Attempts (OR = 0.709, $p < 0.001$) show a negative relationship with the outcome, where each additional submission attempt by the blue fighter decreases the odds of blue winning by 29.1%. This suggests that higher submission attempts by the blue fighter may be linked to a decreased likelihood of blue winning, which may indicate that submission attempts do not effectively contribute to blue's success in this context. For Blue Reach (OR = 0.121, p = 0.006), an increase in blue's reach reduces the odds of red winning. The odds ratio of 0.121 indicates that each unit increase blue's reach significantly lowers the likelihood of red winning, highlighting the importance of reach for blue fighters. In contrast, Red Reach (OR = 7.623, p = 0.0063) increases the odds of red winning by a factor of 7.623 for every unit increase in red's reach, emphasizing the critical role of reach for red fighters in enhancing their chances of victory. Regarding significant strikes, LogBlueSigStr (OR = 0.624, $p < 0.001$) shows that for blue fighters, a higher number of significant strikes landed decreases the odds of red winning. This suggests that effective striking by the blue fighter contributes to their chance of winning by reducing red's odds. Similarly, LogRedSigStr (OR = 1.588, $p < 0.001$) indicates that for red fighters, a higher number of significant strikes landed increases the odds of red winning by 58.8%, reinforcing the importance of striking in determining the fight outcome. Overall, these findings provide insights into the key factors that influence the fight's outcome, with submission attempts, reach, and significant strikes being significant contributors to the likelihood of winning.

```r
# Convert both Red and Blue Submission Attempts back to the original scale
ufc_q2_filtered$actual_LogRedSubAttempts <- exp(ufc_q2_filtered$LogRedSubAttempts) - 1
ufc_q2_filtered$actual_LogBlueSubAttempts <- exp(ufc_q2_filtered$LogBlueSubAttempts) - 1
ufc_q2_filtered$predicted_prob <- predict(final_model, newdata = ufc_q2_filtered, type = "res
# Plot both Red and Blue Submission Attempts against the predicted probabilities
ggplot(ufc_q2_filtered, aes(x = actual_LogRedSubAttempts, y = predicted_prob)) +
    geom_point(aes(color = "Red Submission Attempts"), alpha = 0.1, size = 2) +  # Reduce opa
  geom_smooth(aes(color = "Red Submission Attempts"), method = "loess", size = 1.5) +  # Thic
  geom_point(aes(x = actual_LogBlueSubAttempts, color = "Blue Submission Attempts"), alpha =
  geom_smooth(aes(x = actual_LogBlueSubAttempts, color = "Blue Submission Attempts"), method
  labs(
    title = "Predicted Probability of Red Winning vs. Submission Attempts",
    x = "Submission Attempts",
    y = "Predicted Probability of Red Winning",
    color = "Predictor"
  ) +
  scale_color_manual(
    values = c("Red Submission Attempts" = "red", "Blue Submission Attempts" = "blue")
  ) +  # Adjusted colors for contrast
```

```
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 10),  # Center-align title and increase fo
    axis.title = element_text(size = 10),
    axis.text = element_text(size = 9),
    legend.position = "bottom",  # Move legend to the bottom
    legend.title = element_text(size = 10),
    legend.text = element_text(size = 9)
  )
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```

```
`geom_smooth()` using formula = 'y ~ x'
`geom_smooth()` using formula = 'y ~ x'
```




```
vif_results = vif(final_model)
# Try extracting VIF
```

```r
if (length(vif_results) == 0) {
  stop("The VIF computation returned an empty result. Check your model for singularities or c
} else if (is.matrix(vif_results)) {
  # Handle VIF as a matrix (for categorical predictors)
  vif_table <- data.frame(
    Variable = rownames(vif_results),
    GVIF = vif_results[, 1],
    Df = vif_results[, 2],
    GVIF_Adjusted = vif_results[, 1]^(1 / (2 * vif_results[, 2]))
  )
} else {
  # Handle VIF as a vector (for numeric-only models)
  vif_table <- data.frame(
    Variable = names(vif_results),
    VIF = vif_results
  )
}

# Add descriptive names (optional)
vif_table$Variable <- c(
  "Log Red Submission Attempts",
  "Log Blue Submission Attempts",
  "Log Blue Reach",
  "Log Red Reach",
  "Log Blue Significant Strikes",
  "Log Red Significant Strikes"
)

# Remove row names
rownames(vif_table) <- NULL

kable(
  vif_table,
  caption = "Variance Inflation Factor (VIF) and Adjusted GVIF",
  align = "c",
  digits = 3
)
```

Table 4: Variance Inflation Factor (VIF) and Adjusted GVIF

Variable	VIF
Log Red Submission Attempts	1.043
Log Blue Submission Attempts	1.036
Log Blue Reach	2.183
Log Red Reach	2.183
Log Blue Significant Strikes	3.868
Log Red Significant Strikes	3.898

```
print(vif_results)
```

```
 LogRedSubAttempts LogBlueSubAttempts       LogBlueReach        LogRedReach 
          1.042530           1.035591           2.183275           2.183372 
      LogBlueSigStr        LogRedSigStr 
          3.867791           3.897926 
```

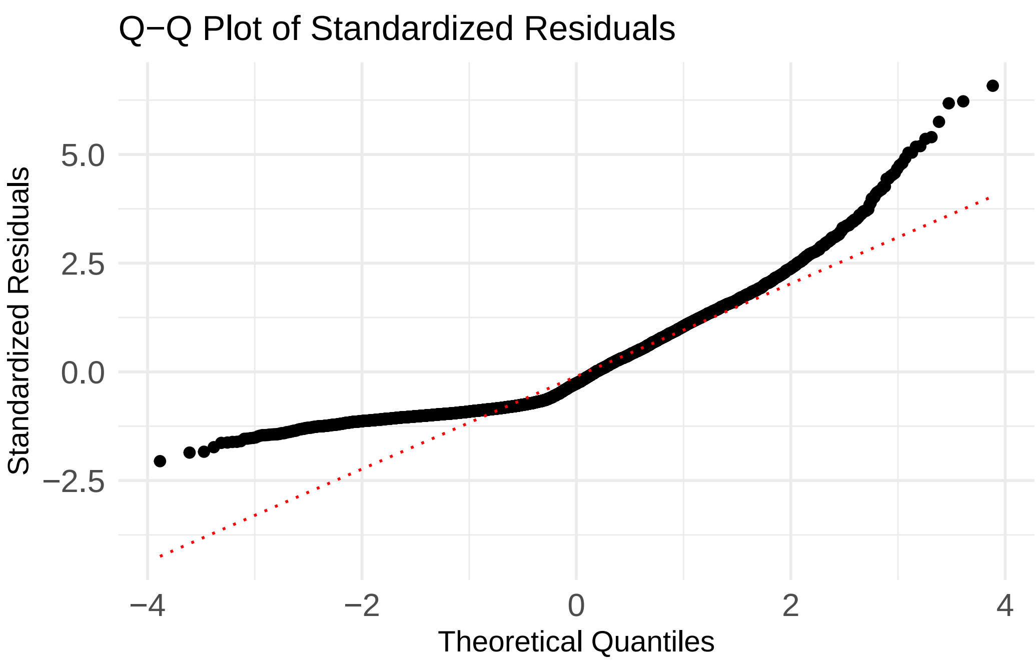
```
vif_results = vif(final_model)

kable(vif_results, caption = "Variance Inflation Factor (VIF)", align = "c", digits = 3)
```

Table 5: Variance Inflation Factor (VIF)

	x
LogRedSubAttempts	1.043
LogBlueSubAttempts	1.036
LogBlueReach	2.183
LogRedReach	2.183
LogBlueSigStr	3.868
LogRedSigStr	3.898

Appendix

1. Linearity Assumption for the initial model for research question 1



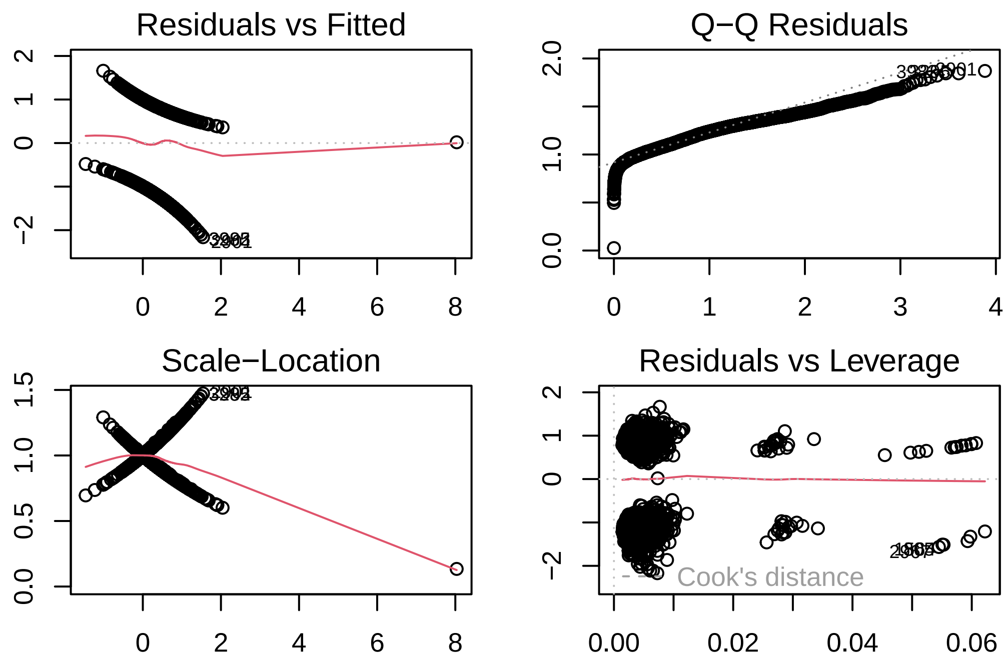
2. Research Question 2

```r
# Model summary of the initial simple logistic model
summary(sim_logistic_model)
```

```
Call:
glm(formula = Outcome ~ LogRedSubAttempts + LogBlueSubAttempts +
    LogBlueReach + LogRedReach + LogBlueSigStr + LogRedSigStr +
    LogFightTime + WeightClass, family = binomial, data = ufc_q2)

Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                      -5.351057   6.459545  -0.828 0.407447
LogRedSubAttempts                 0.405680   0.089319   4.542 5.57e-06 ***
LogBlueSubAttempts               -0.292197   0.084704  -3.450 0.000561 ***
LogBlueReach                     -1.439347   0.907894  -1.585 0.112883
LogRedReach                       2.491817   0.903558   2.758 0.005819 **
LogBlueSigStr                    -0.365003   0.048374  -7.545 4.51e-14 ***
LogRedSigStr                      0.373853   0.050081   7.465 8.33e-14 ***
LogFightTime                      0.035227   0.032697   1.077 0.281313
WeightClassCatch Weight          -0.057436   0.343726  -0.167 0.867293
WeightClassFeatherweight         -0.027917   0.131144  -0.213 0.831425
```

```
WeightClassFlyweight              -0.003566   0.159466  -0.022 0.982161
WeightClassHeavyweight            -0.045397   0.208875  -0.217 0.827942
WeightClassLight Heavyweight      -0.176252   0.192860  -0.914 0.360776
WeightClassLightweight            -0.136914   0.126242  -1.085 0.278128
WeightClassMiddleweight           -0.276099   0.164307  -1.680 0.092883 .
WeightClassWelterweight           -0.281531   0.145431  -1.936 0.052888 .
WeightClassWomen's Bantamweight  -0.177841   0.191071  -0.931 0.351979
WeightClassWomen's Featherweight  0.230400   0.512452   0.450 0.652996
WeightClassWomen's Flyweight     -0.122585   0.184749  -0.664 0.506998
WeightClassWomen's Strawweight    0.011955   0.187268   0.064 0.949100
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6674.5  on 4894  degrees of freedom
Residual deviance: 6566.3  on 4875  degrees of freedom
AIC: 6606.3

Number of Fisher Scoring iterations: 4
```

```r
par(mfrow=c(2,2), mar = c(2,2,2,2))

plot(sim_logistic_model)
```




```
# Model summary of the step model
summary(step_model)
```

```
Call:
glm(formula = Outcome ~ LogRedSubAttempts + LogBlueSubAttempts +
    LogBlueReach + LogRedReach + LogBlueSigStr + LogRedSigStr,
    family = binomial, data = ufc_q2)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)         0.36477    2.73096   0.134 0.893744
LogRedSubAttempts   0.39057    0.08768   4.454 8.41e-06 ***
LogBlueSubAttempts -0.29815    0.08385  -3.556 0.000377 ***
LogBlueReach       -1.90853    0.71404  -2.673 0.007521 **
LogRedReach         1.88597    0.69892   2.698 0.006967 **
LogBlueSigStr      -0.36729    0.04821  -7.618 2.57e-14 ***
LogRedSigStr        0.36873    0.04979   7.405 1.31e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
```

```
    Null deviance: 6674.5  on 4894  degrees of freedom
Residual deviance: 6577.8  on 4888  degrees of freedom
AIC: 6591.8

Number of Fisher Scoring iterations: 4
```

```r
par(mfrow=c(2,2), mar = c(2,2,2,2))

plot(step_model)
```