

Final Project

Yifei Yang, Yiming Yuan

Read in the data

```
library(skimr)
library(ggplot2)
library(dplyr)
library(readr)
library(MASS)

ufc <- read_csv("data/ufc-fighters-statistics.csv")
```

Introduction and data

Around 300 million individuals worldwide identify themselves as fans of Mixed Martial Arts (MMA), with its popularity peaking in nations such as the United States, the United Kingdom, Brazil, Singapore, and China. The Ultimate Fighting Championship (UFC) is the premier organization in the MMA world. Our motivation is to research on what contributes to the fighters' performance. Thus, the research questions is: Which model we consider effective in predicting fighters' winning ratio? How fighters improve their performance?

Today's data are UFC fighter statistics including:

wins: The number of wins the fighter has in their career.

draws: The number of draws the fighter has in their career.

height_cm: The height of the fighter in centimeters.

weight_in_kg: The weight of the fighter in kilograms.

reach_in_cm: The reach of the fighter in centimeters.

stance: The fighting stance of the fighter (Orthodox/Southpaw/Switch).

significant_strikes_landed_per_minute: The average number of significant strikes landed by the fighter per minute.

significant_striking_accuracy: The percentage of significant strikes that land successfully for the fighter.

significant_strikes_absorbed_per_minute: The average number of significant strikes absorbed by the fighter per minute.

significant_strike_defence: The percentage of opponent's significant strikes that the fighter successfully defends.

average_takedowns_landed_per_15_minutes: The average number of takedowns landed by the fighter per 15 minutes.

takedown_accuracy: The percentage of takedown attempts that are successful for the fighter.

takedown_defense: The percentage of opponent's takedown attempts that the fighter successfully defends.

average_submissions_attempted_per_15_minutes: The average number of submission attempts made by the fighter per 15 minutes.

Our response variable is wins, predictors are

sources:

<https://www.kaggle.com/datasets/aaronfriasr/ufc-fighters-statistics?resource=download>

<https://www.euronews.com/business/2023/09/27/the-booming-billion-dollar-business-of-combat-sports>

Data Preparation

```
# A tibble: 6 x 20
  name      nickname wins losses draws height_cm weight_in_kg reach_in_cm stance
<chr>    <chr>    <dbl> <dbl> <dbl>    <dbl>      <dbl>      <dbl> <chr>
1 Robert ~ <NA>      7      0      0     190.      93.0        NA Ortho~
2 Daniel ~ The Ani~ 15     37      0     185.      83.9        NA <NA>
3 Dan Mol~ <NA>     13      9      0     178.      98.0        NA <NA>
4 Paul Ru~ <NA>      7      4      0     168.      61.2        NA <NA>
5 Collin ~ All In    8      2      0     190.      83.9       193. Ortho~
6 Gerald ~ The Fin~  9      7      0     175.      70.3        NA Ortho~
# i 11 more variables: date_of_birth <date>,
#   significant_strikes_landed_per_minute <dbl>,
#   significant_striking_accuracy <dbl>,
```

```
# significant_strikes_absorbed_per_minute <dbl>,
# significant_strike_defence <dbl>,
# average_takedowns_landed_per_15_minutes <dbl>, takedown_accuracy <dbl>,
# takedown_defense <dbl>, ...
```

Table 1: Data summary

| | |
|------------------------|------|
| Name | ufc |
| Number of rows | 4111 |
| Number of columns | 20 |
| Column type frequency: | |
| character | 3 |
| Date | 1 |
| numeric | 16 |
| Group variables | None |

Variable type: character

| skim_variable | n_missing | complete_rate | min | max | empty | n_unique | whitespace |
|---------------|-----------|---------------|-----|-----|-------|----------|------------|
| name | 0 | 1.00 | 5 | 27 | 0 | 4105 | 0 |
| nickname | 1854 | 0.55 | 1 | 30 | 0 | 1784 | 0 |
| stance | 823 | 0.80 | 6 | 11 | 0 | 5 | 0 |

Variable type: Date

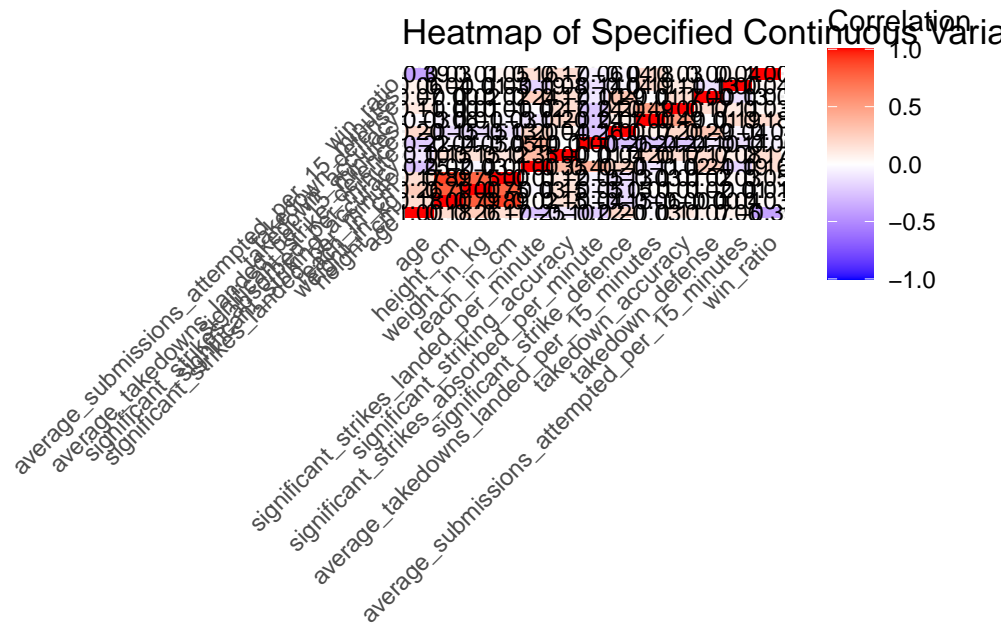
| skim_variable | n_missing | complete_rate | min | max | median | n_unique |
|---------------|-----------|---------------|------------|------------|------------|----------|
| date_of_birth | 1135 | 0.72 | 1943-01-25 | 2004-10-08 | 1986-11-06 | 2565 |

Variable type: numeric

| skim_variable | n_missing | complete_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
|---------------|-----------|---------------|--------|-------|--------|--------|--------|--------|--------|------|
| wins | 0 | 1.00 | 12.37 | 9.37 | 0.00 | 7.00 | 11.00 | 17.00 | 253.00 | |
| losses | 0 | 1.00 | 5.73 | 5.10 | 0.00 | 2.00 | 5.00 | 8.00 | 83.00 | |
| draws | 0 | 1.00 | 0.26 | 0.82 | 0.00 | 0.00 | 0.00 | 0.00 | 11.00 | |
| height_cm | 298 | 0.93 | 178.23 | 8.89 | 152.40 | 172.72 | 177.80 | 185.42 | 226.06 | |
| weight_in_kg | 87 | 0.98 | 77.40 | 17.98 | 47.63 | 65.77 | 77.11 | 83.91 | 349.27 | |

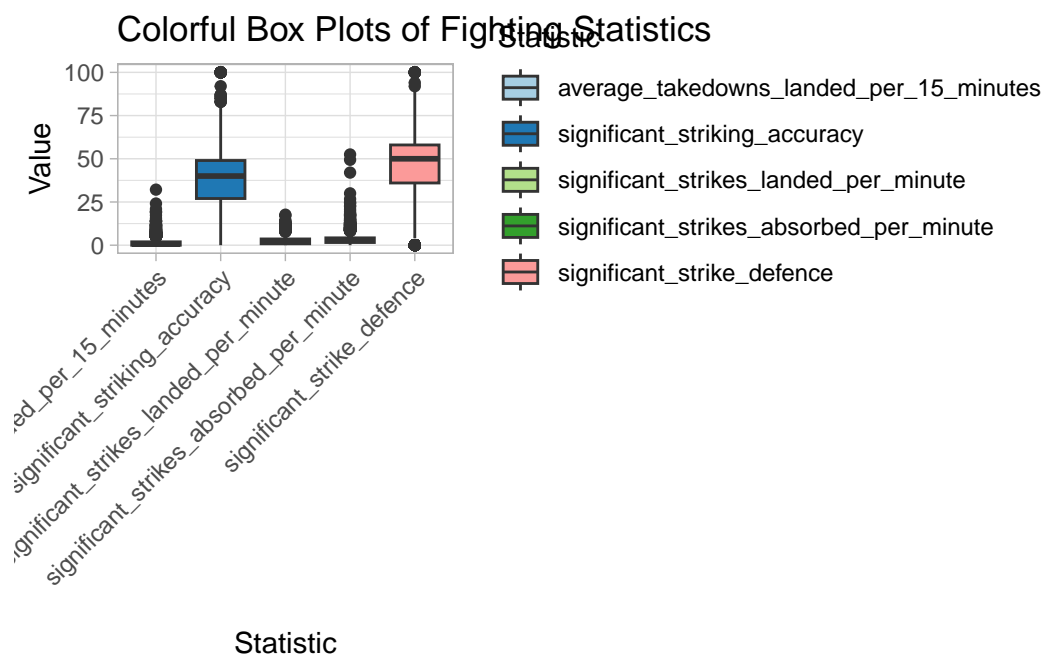
| skim_variable | n_missing | complete | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
|--|-----------|----------|--------|-------|--------|--------|--------|--------|--------|------|
| reach_in_cm | 1927 | 0.53 | 181.81 | 10.68 | 147.32 | 175.26 | 182.88 | 190.50 | 213.36 | |
| significant_strikes_landed_per_minute | 0 | 1.00 | 2.44 | 1.99 | 0.00 | 0.83 | 2.33 | 3.60 | 17.65 | |
| significant_striking_accuracy | 0 | 1.00 | 35.54 | 20.40 | 0.00 | 27.00 | 40.00 | 49.00 | 100.00 | |
| significant_strikes_absorbed_per_minute | 0 | 1.00 | 3.15 | 2.85 | 0.00 | 1.55 | 2.94 | 4.23 | 52.50 | |
| significant_strike_defence | 0 | 1.00 | 42.64 | 22.32 | 0.00 | 36.00 | 50.00 | 58.00 | 100.00 | |
| average_takedowns_landed_per_15_minutes | 0 | 1.00 | 1.25 | 1.94 | 0.00 | 0.00 | 0.59 | 1.94 | 32.14 | |
| takedown_accuracy | 0 | 1.00 | 26.30 | 28.70 | 0.00 | 0.00 | 22.00 | 45.00 | 100.00 | |
| takedown_defense | 0 | 1.00 | 38.96 | 34.43 | 0.00 | 0.00 | 42.00 | 66.00 | 100.00 | |
| average_submissions_attempted_per_15_minutes | 0 | 1.00 | 0.61 | 1.51 | 0.00 | 0.00 | 0.00 | 0.70 | 21.90 | |
| win_ratio | 19 | 1.00 | 0.66 | 0.19 | 0.00 | 0.60 | 0.69 | 0.78 | 1.00 | |
| age | 1135 | 0.72 | 38.59 | 7.78 | 20.00 | 33.00 | 38.00 | 44.00 | 81.00 | |

Choose Predictors



Predictors: age Stance (category) average_takedowns_landed_per_15_minutes Significant_striking_accuracy Significant_strikes_landed_per_minute significant_strikes_absorbed_per_minute significant_strike_defence

EDA



Model Fitting

Call:

```
lm(formula = win_ratio ~ age + stance + average_takedowns_landed_per_15_minutes +
    significant_striking_accuracy + significant_strikes_landed_per_minute +
    significant_strikes_absorbed_per_minute + significant_strike_defence,
    data = ufc)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|----------|----------|---------|---------|---------|
| -0.70326 | -0.06308 | 0.00335 | 0.07501 | 0.58006 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|---|------------|------------|---------|----------|
| (Intercept) | 0.9418870 | 0.0539846 | 17.447 | < 2e-16 |
| age | -0.0082692 | 0.0003249 | -25.453 | < 2e-16 |
| stanceOrthodox | 0.0179482 | 0.0502243 | 0.357 | 0.721 |
| stanceSideways | 0.1840189 | 0.1322619 | 1.391 | 0.164 |
| stanceSouthpaw | 0.0236782 | 0.0504429 | 0.469 | 0.639 |
| stanceSwitch | 0.0238210 | 0.0510905 | 0.466 | 0.641 |
| average_takedowns_landed_per_15_minutes | 0.0095566 | 0.0013340 | 7.164 | 9.86e-13 |

| | | | | |
|---|------------|-----------|--------|----------|
| significant_striking_accuracy | 0.0009264 | 0.0001842 | 5.029 | 5.22e-07 |
| significant_strikes_landed_per_minute | 0.0063530 | 0.0016135 | 3.937 | 8.42e-05 |
| significant_strikes_absorbed_per_minute | -0.0051719 | 0.0009167 | -5.642 | 1.84e-08 |
| significant_strike_defence | 0.0001415 | 0.0001732 | 0.817 | 0.414 |

(Intercept) ***

age ***

stanceOrthodox

stanceSideways

stanceSouthpaw

stanceSwitch

average_takedowns_landed_per_15_minutes ***

significant_striking_accuracy ***

significant_strikes_landed_per_minute ***

significant_strikes_absorbed_per_minute ***

significant_strike_defence

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

Residual standard error: 0.1224 on 2965 degrees of freedom

(1135 observations deleted due to missingness)

Multiple R-squared: 0.2876, Adjusted R-squared: 0.2852

F-statistic: 119.7 on 10 and 2965 DF, p-value: < 2.2e-16

Categorical variable:

0.0435211 represents the difference in average win ratio for fighters with Orthodox stance, compared with fighters with Open stance, while holding the other predictors constant.

-0.0893852 represents the difference in average win ratio for fighters with Sideways stance, compared with fighters with Open stance, while holding the other predictors constant.

0.0484787 represents the difference in average win ratio for fighters with Southpaw stance, compared with fighters with Open stance, while holding the other predictors constant.

0.0895293 represents the difference in average win ratio for fighters with Switch stance, compared with fighters with Open stance, while holding the other predictors constant.

We notice that the p-value for Stance variables are all less than 0.05, and thus, there is insufficient evidence to suggest a linear relationship between fighter's stance and wins ratio at 0.05 significance level, while controlling for other predictors

Continuous variable:

While holding the other predictors constant, when the average number of takedowns landed by the fighter per 15 minutes increases by 1 time, the average win ratio of the fighter will increase by 0.0113384.

While holding the other predictors constant, when the significant striking accuracy increases by 1%, the average win ratio of the fighter will increase by 0.0017843.

While holding the other predictors constant, when the average number of significant strikes landed by the fighter per minute increases by 1 time, the average win ratio of the fighter will increase by 0.0199915.

While holding the other predictors constant, when the average number of significant strikes absorbed by the fighter per minute increases by 1 time, the average win ratio of the fighter will decrease by 0.0016777.

While holding the other predictors constant, when the opponent's significant strikes that the fighter successfully defends increases by 1%, the average win ratio of the fighter will increase by 0.0009887.

Call:

```
polr(formula = win_ratio_group ~ age + stance + average_takedowns_landed_per_15_minutes +
      significant_striking_accuracy + significant_strikes_landed_per_minute +
      significant_strikes_absorbed_per_minute + significant_strike_defence,
      data = ufc)
```

Coefficients:

| | Value | Std. Error | t value |
|---|------------|------------|----------|
| age | -0.1151447 | 0.006670 | -17.2641 |
| stanceOrthodox | -0.3937706 | 0.816083 | -0.4825 |
| stanceSideways | 4.3460381 | 4.460038 | 0.9744 |
| stanceSouthpaw | -0.1067159 | 0.821314 | -0.1299 |
| stanceSwitch | -0.5320075 | 0.838556 | -0.6344 |
| average_takedowns_landed_per_15_minutes | 0.1553674 | 0.031796 | 4.8864 |
| significant_striking_accuracy | 0.0117169 | 0.003527 | 3.3222 |
| significant_strikes_landed_per_minute | 0.0675558 | 0.034174 | 1.9768 |
| significant_strikes_absorbed_per_minute | -0.0528466 | 0.018069 | -2.9248 |
| significant_strike_defence | -0.0007179 | 0.003301 | -0.2175 |

Intercepts:

| | Value | Std. Error | t value |
|-------------|---------|------------|---------|
| Low Medium | -8.6075 | 0.9195 | -9.3612 |
| Medium High | -5.0338 | 0.9050 | -5.5622 |

Residual Deviance: 3562.268

AIC: 3586.268
 (1135 observations deleted due to missingness)

| Actual | Predicted | | |
|--------|-----------|--------|------|
| | Low | Medium | High |
| Low | 2 | 47 | 7 |
| Medium | 0 | 251 | 620 |
| High | 0 | 191 | 1858 |

[1] "Accuracy: 0.709341397849462"

Call:

```
lm(formula = win_ratio ~ age + stance + average_takedowns_landed_per_15_minutes +
    significant_striking_accuracy + significant_strikes_landed_per_minute +
    significant_strikes_absorbed_per_minute + significant_strike_defence +
    age * average_takedowns_landed_per_15_minutes + age * significant_striking_accuracy +
    age * significant_strikes_absorbed_per_minute + age * significant_strike_defence,
    data = ufc)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|----------|----------|---------|---------|---------|
| -0.72177 | -0.06487 | 0.00156 | 0.07423 | 0.66218 |

Coefficients:

| | Estimate | Std. Error | t value |
|---|------------|------------|---------|
| (Intercept) | 1.130e+00 | 6.512e-02 | 17.358 |
| age | -1.218e-02 | 8.557e-04 | -14.230 |
| stanceOrthodox | 5.375e-03 | 4.994e-02 | 0.108 |
| stanceSideways | 1.853e-01 | 1.315e-01 | 1.410 |
| stanceSouthpaw | 9.690e-03 | 5.017e-02 | 0.193 |
| stanceSwitch | 1.104e-02 | 5.080e-02 | 0.217 |
| average_takedowns_landed_per_15_minutes | -3.342e-03 | 6.007e-03 | -0.556 |
| significant_striking_accuracy | -1.792e-03 | 7.956e-04 | -2.253 |
| significant_strikes_landed_per_minute | 7.883e-03 | 1.656e-03 | 4.760 |
| significant_strikes_absorbed_per_minute | 5.719e-03 | 4.378e-03 | 1.306 |
| significant_strike_defence | -1.763e-03 | 7.565e-04 | -2.331 |
| age:average_takedowns_landed_per_15_minutes | 3.698e-04 | 1.615e-04 | 2.289 |
| age:significant_striking_accuracy | 6.309e-05 | 1.885e-05 | 3.347 |
| age:significant_strikes_absorbed_per_minute | -3.119e-04 | 1.141e-04 | -2.734 |
| age:significant_strike_defence | 4.203e-05 | 1.773e-05 | 2.370 |

Pr(>|t|)


```

(Intercept)                < 2e-16 ***
age                        < 2e-16 ***
stanceOrthodox             0.914298
stanceSideways             0.158761
stanceSouthpaw             0.846868
stanceSwitch               0.827888
average_takedowns_landed_per_15_minutes 0.578025
significant_striking_accuracy 0.024348 *
significant_strikes_landed_per_minute 2.03e-06 ***
significant_strikes_absorbed_per_minute 0.191499
significant_strike_defence 0.019831 *
age:average_takedowns_landed_per_15_minutes 0.022133 *
age:significant_striking_accuracy 0.000826 ***
age:significant_strikes_absorbed_per_minute 0.006295 **
age:significant_strike_defence 0.017840 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1215 on 2961 degrees of freedom
(1135 observations deleted due to missingness)
Multiple R-squared:  0.2995,    Adjusted R-squared:  0.2962
F-statistic: 90.41 on 14 and 2961 DF,  p-value: < 2.2e-16

```