# Examining factors that may affect the time at which a finish occurs during a UFC contest scheduled for 3 rounds.

## INTRODUCTION

Mixed Martial Arts (MMA) is a relatively new combat sport. Every MMA contest (or bout) is contested between 2 trained athletes of the same weight, who battle for 3, 5-minute rounds, or 5, 5-minute rounds. The contestants are allowed to use any combination of kickboxing, freestyle wrestling, and submission grappling in order to win the contest with one of the possible outcomes being a stoppage (Knockout, Technical Knockout, or submission), doctor's stoppage due to injury, tap-out when one of the contestants decides they no longer want to continue, disqualification in case of a foul, or referee/corner stoppage when the referee/coaches deem the bout needs to be stopped in order to protect the contestant (What is UFC, what is MMA and what are the rules?, 2017).

Most of the research done on MMA is focused on understanding and preventing injuries that usually occur during training or competition (Rainey C. E. (2009)), determining MMA specific training by using science as a way to optimize the benefits of training, and insuring optimal preparation for the contest (Bounty, P. L., Campbell, B. I., Galvan, E., Cooke, M., & Antonio, J. (2011)). A small part of the research examines the way with which a contest ends (Buse, G. J. (2006)), or how the tactical approach or the contestants affects the outcome of the bout (Miarka, B., Munoz, A. E., Perez, D. I. V., Teixeira, F. G., Brito, C. J. (2020)).

The goal of this study is to understand how athletes' body characteristics and prior high level contest experience affects how fast a 3-round MMA contest under the UFC organization ends with a stoppage due to KO/TKO or a Submission (i.e., if it makes it occur earlier, or later. We do not examine the particular time of finish, but rather if that time increases or decreases given the change in specific factors). This is an original question that aims to provide further understanding of the factors that affect how fast one trained athlete overpowers and finishes another trained athlete, and thus would constitute a welcome addition to the literature. We expect higher discrepancy in skill level and physical attributes (longer arms, or taller in height), as well as weight class to result in a faster finish, as MMA is a highly tactical and physically demanding contact sport.

## METHODS

We select the variables that seem important. These variables must thus capture the discrepancy between physical attributes, and skill level of the contestants.

We need to do the PDF and PMF of each variable in order to identify problematic trends, as well as do the scatterplot/boxplot of each variable with the response in order to observe any nonlinear relationships, fanning, or problematic observations.

We will fit N simple models for the N predictors to observe how strong the relationship between each chosen predictor and the response is.

We will then fit the full model and check its Y vs Yhat plot (our data should be close to the identity function) and its pairwise predictor plot (they should display linear or random patterns). If these conditions are satisfied, we will plot the residuals vs fitted values plot, and the normal QQ-plot. If the data is not as normal as we'd like, or if there is fanning, we will apply a BOX-COX transformation and then recheck the plots stated here to see if we have desired results.

Then we will reduce the model by removing one by one the predictors which had initially high p-values, each time we will fit a new model, do ANOVA F-TEST with the previous model, and assuming it has passed the test, we check the new model for the same conditions and assumptions as after the BOX-COX.

We will classify problematic observations using hii (leverage points), cooks distance, DFBETAS, DFFITS (influential observations) and standardized residuals (outliers)

We will determine if the predictors have multicollinearity by estimating the VIF of our model.

Lastly, we will use the summary and the 95% CIs of the predictors in order to observe how fluctuations in each predictor fluctuates the mean response.

**RESULTS**

Numeric variables: ContestDuration, ReachDifference, AgeDifference, TitleExperienceDif

Categorical Variables: WeightClass, Year, FinishType

We plotted the PDFs and PMFs of all variables below. Sample size = 1361, 7 variables.

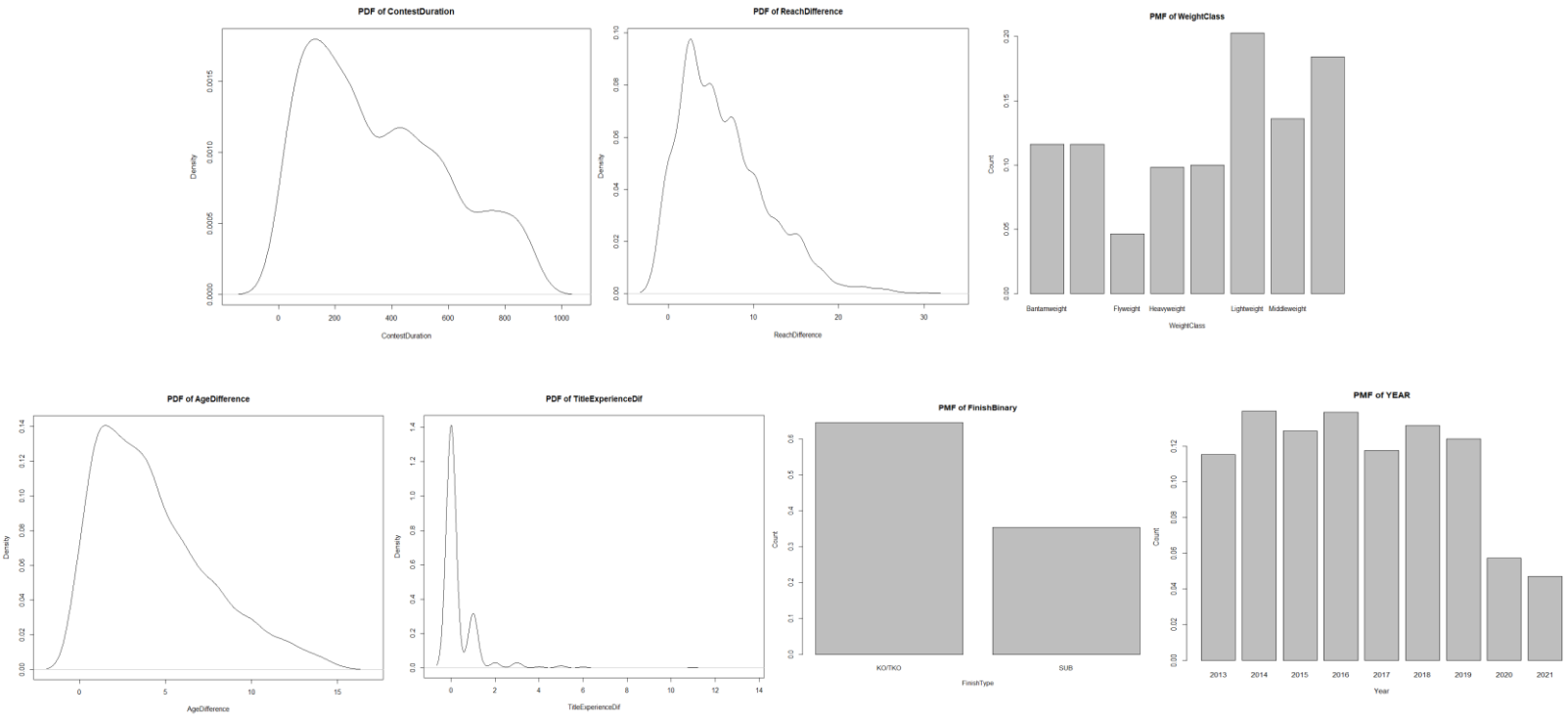



Figure 1: Probability distributions of all variables

We can clearly observe a right skew in the response, as well as in the rest of the variables apart from Year. Noting possible need for transformation in these skewed variables. Also, in the scatterplots (Appendix: P1) note large spread of data, and datapoints far away from the center spread.

Fitting the model:

**ContestDuration ~ AgeDifference + ReachDifference + TitleExperienceDif + WeightClass + FinishType + Year**

Checking the regression conditions and assumptions:

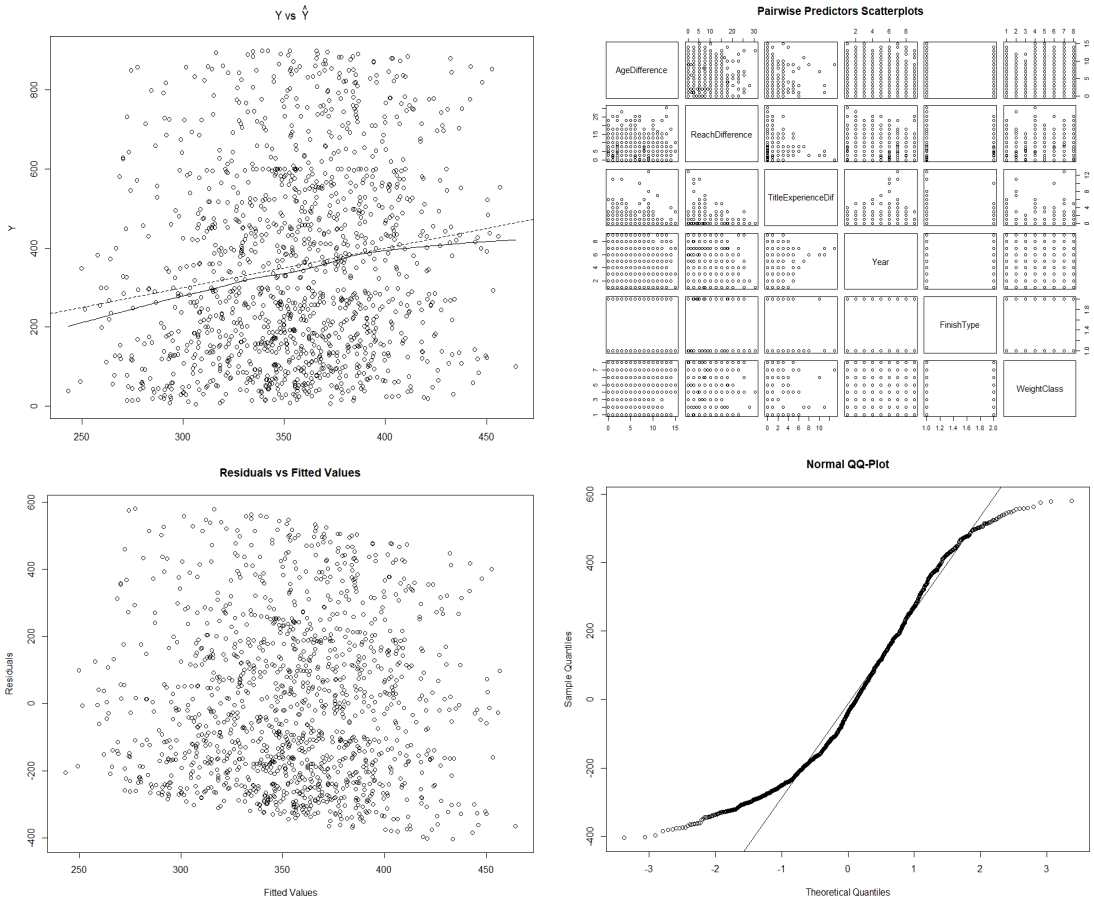



Figure 2: Regression Conditions (top 2 plots) Regression Assumptions (bottom 2 plots)

We can see that the pairwise predictors plot displays either random or linear patterns, while in the Y vs Yhat plot, our data seems to be quite close to the identity function. The two conditions are thus satisfied, and we can use the Residuals vs Fitted values, and Normal QQ-Plots in order to understand how we can improve our model. In the residuals vs fitted values plot, it seems that the residuals are randomly spread, and only minimal fanning is noted, meaning that the constant variance assumption is satisfied. In the normal QQ-plot, we notice significant deviations away from the identity function at the edges, and that the points do not adhere perfectly to the identity

in the middle. We will thus go ahead with the BOX-COX transformation in order to improve the normality assumption in our model.

We conduct BOX-COX only for the numeric variables ReachDifference, AgeDifference and ContestDuration. The resulting model is:

**ContestDuration^(1/2) ~ AgeDifference^(0.06) + ReachDifference^(0.04) + TitleExperienceDif + WeightClass + FinishType + Year**

In this transformation, we had to add the number 10^(-100) to the AgeDifference and ReachDifference variables, as the BOX-COX transformation fails when given 0 values. Thus, adding such a miniscule number to the data does not practically change their values, but allows the BOX-COX transformation to work. We again check the Conditions and Assumptions of the transformed model looking at the same plots as above.

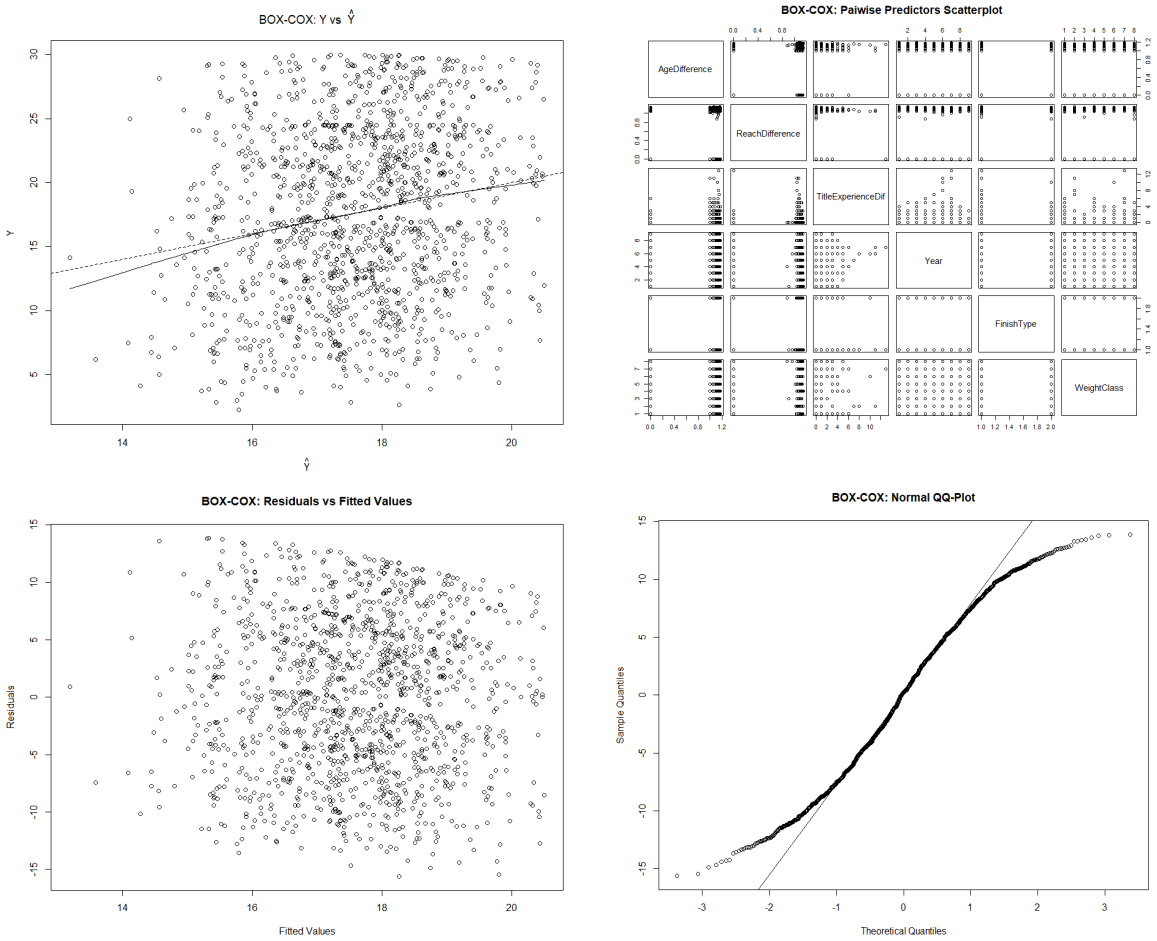



Figure 3:Transformed Model Regression Conditions (top 2 plots) Regression Assumptions (bottom 2 plots)

The conditions are again satisfied, and we note a slight improvement of normality (variance is same as in the initial model). BOX-COX transformation yielded small but needed improvement in normality.

The predictors with the highest p-values when fitting the full model initially are AgeDifference (highest), TitleExperienceDifference (second highest). The model passed the ANOVA F-TEST at 90% level only when we removed the AgeDifference variable. It failed this test when removing any other high p-value predictor. The resulting model is:

**ContestDuration^(1/2) ~ ReachDifference^(0.04) + TitleExperienceDif + WeightClass + FinishType + Year**

This model was also found to satisfy the regression Conditions and Assumptions by the same process as before.

It was found that 378 of the 1321 datapoints are influential points, 5 of which are also leverage points (21 total leverage points) (see Appendix: P2).

The VIF of each predictor was less than 1.2, well within the accepted range of (-5,5).

Below we present the 95% CIs and the estimates for all the coefficients

| Variable | 2.50% | 97.50% | Estimate | P-value |
|---|---|---|---|---|
| (Intercept) | 14.3652708 | 18.06446963 | 16.21487023 | 4.64E-60 |
| ReachDifference | 0.2167129 | 2.29096146 | 1.25383718 | 0.02 |
| WeightClassFeatherweight | -1.9528701 | 1.04495669 | -0.45395669 | 0.55 |
| WeightClassFlyweight | -1.8208084 | 2.15448408 | 0.16683786 | 0.87 |
| WeightClassHeavyweight | -3.6491316 | -0.48667299 | -2.0679023 | 0.01 |
| WeightClassLight Heavyweight | -4.1113918 | -0.97551242 | -2.54345213 | 0.0015 |
| WeightClassLightweight | -2.0863627 | 0.57991904 | -0.75322185 | 0.27 |
| WeightClassMiddleweight | -2.0296975 | 0.86944245 | -0.58012754 | 0.43 |
| WeightClassWelterweight | -2.644511 | 0.07975257 | -1.28237921 | 0.065 |
| TitleExperienceDif | -0.2736691 | 0.43877739 | 0.08255417 | 0.65 |
| FinishTypeSUB | 0.2514927 | 1.78709252 | 1.01929259 | 0.0093 |
| Year2014 | -1.0868792 | 1.78838327 | 0.35075203 | 0.63 |
| Year2015 | -1.0868806 | 1.84298513 | 0.37805228 | 0.61 |
| Year2016 | 0.2277859 | 3.11377497 | 1.67078042 | 0.02 |
| Year2017 | -0.4599414 | 2.53370321 | 1.03688091 | 0.17 |
| Year2018 | 0.2270107 | 3.14984421 | 1.68842745 | 0.02 |
| Year2019 | -1.2203039 | 1.74359363 | 0.26164485 | 0.73 |
| Year2020 | -1.901515 | 1.78955333 | -0.05598085 | 0.95 |
| Year2021 | -0.5709682 | 3.38731061 | 1.40817119 | 0.16 |

Table 1: Estimates and 95% Confidence Intervals for all the predictors in the final model

## DISCUSSION

**Given that all the other predictors are held fixed, and the only changes occur in the predictor WeightClass:**

The categories: Featherweight, Flyweight, Lightweight, Middleweight and Welterweight all have p-values > 0.05 compared to the baseline category of bantamweight, thus we consider them to not be significantly different from the baseline category (Bantamweight) with respect to their effect on the time a finish occurs. In contrast, we see that the weightclasses Heavyweight and Lightheavyweight have p-values < 0.05, which means that we consider them to be significantly different from the baseline category with respect to their effect on the square root of the response. In particular, because the upper bound of their 95% confidence interval is negative, we interpret this as a finish occurring on average earlier for bouts contested between Lightheavyweights, and between Heavyweights, compared to a finish occurring during a bout between Bantamweights.

**Given that all the other predictors are held fixed, and the only changes occur in the predictor FinishType:**

The estimate of the coefficient of FinishTypeSUB has a p-value < 0.05, which means that we consider a finish by submission to be significantly different from the baseline category (Finish by KO/TKO) with respect to its effect on the square root of the response. In particular, because the lower bound of the estimate of this category is positive, we interpret this as a finish by submission occurring on average later than a finish by KO/TKO.

**Given that all the other predictors are held fixed, and the only changes occur in the predictor ReachDifference:**

The estimate of the coefficient of ReachDifference has a p-value < 0.05. This p-value though is for the variable ReachDifference^(0.04), thus we cannot comment on how much exactly the mean response would change by given a unit increase in this variable. The lower bound of this estimate's confidence interval is positive though, so we can interpret this as: bouts between contestants with larger differences in arm reach, have a finish later than bouts between contestants with smaller differences in arm reach.

**Given that all the other predictors are held fixed, and the only changes occur in the predictor Year:**

Given the baseline category of year (2013) the only categories that have p-values < 0.05 are 2018 and 2016. We can thus consider these two years to be significantly different from the baseline category with respect to their effect on the square root of the response. While the rest of the years do not seem to be different from the baseline category with respect to their effect on the response (they have p-values > 0.05). Furthermore, given that for 2018 and 2016 the lower

bound of their 95% confidence intervals is positive, we can say that for these two years, the bouts on average saw a finish later than the year 2013.

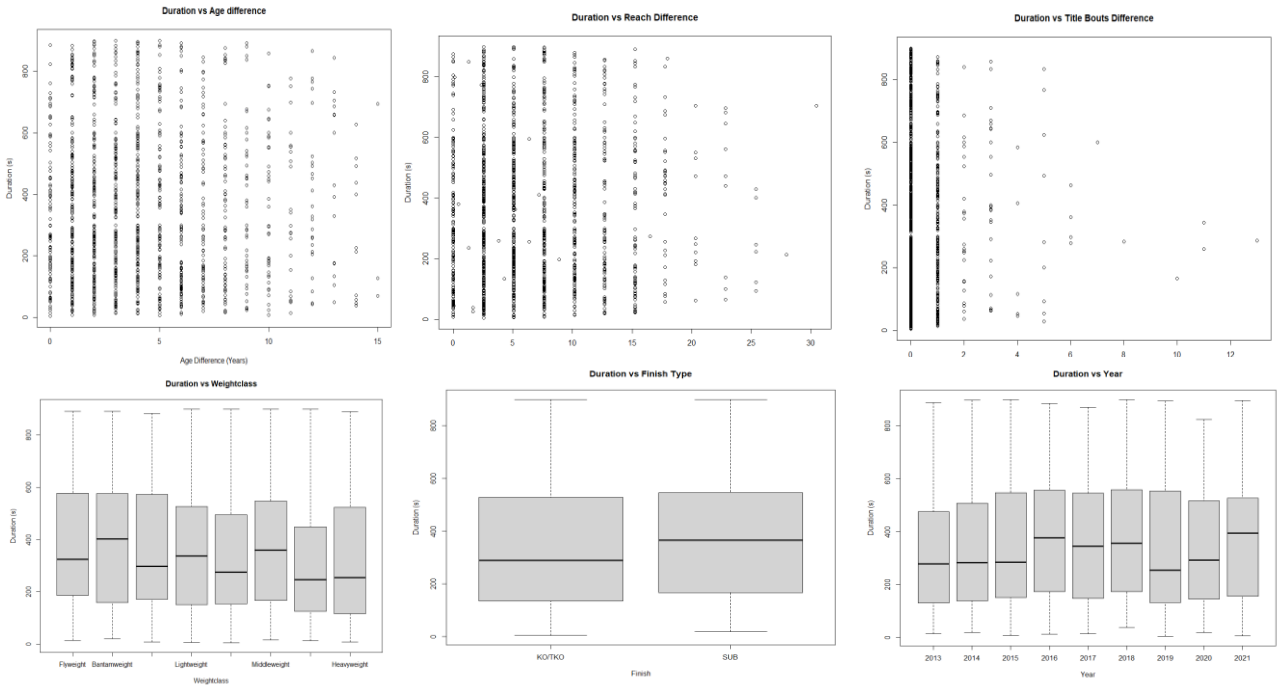
**Given that all the other predictors are held fixed, and the only changes occur in the predictor TitleExperienceDif:**

The estimate of the coefficient of this predictor has p-value > 0.05, thus we cannot say that a unit change in this predictor causes a change in the square root of the mean response.

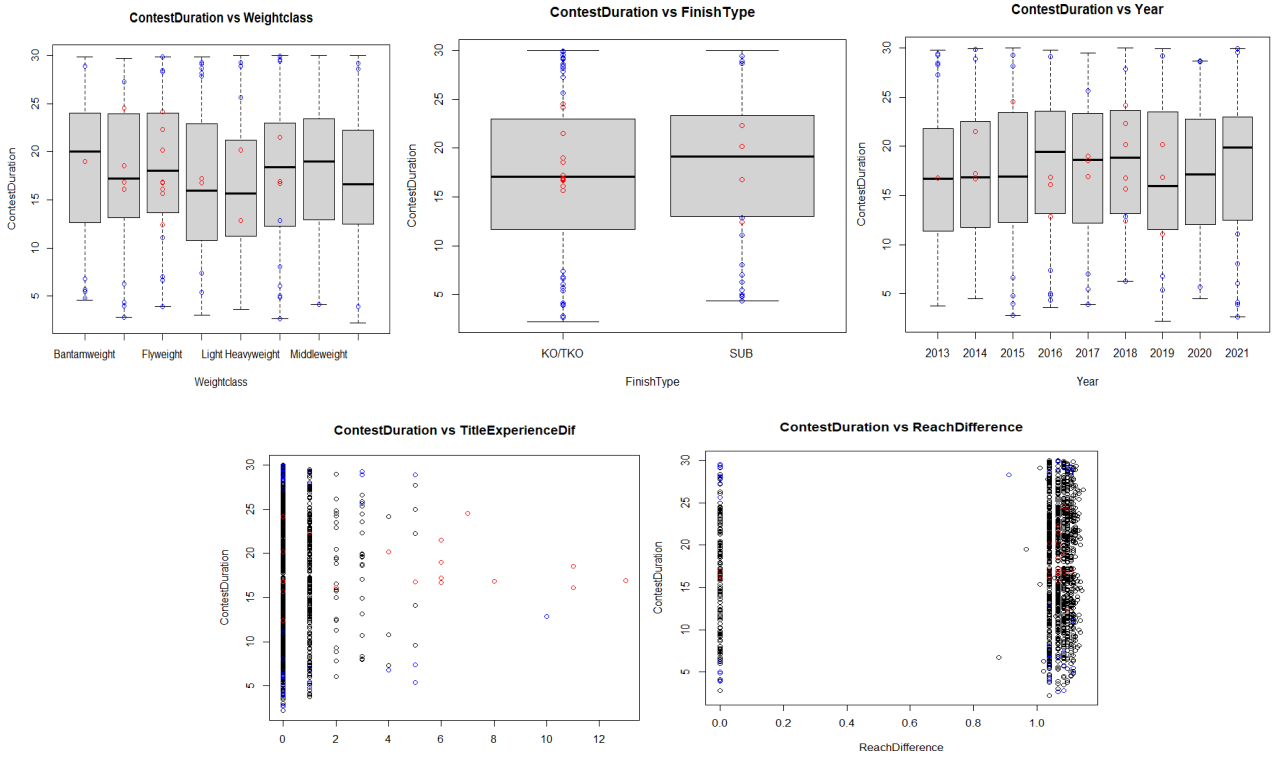
Given the above explanation of our model, we can conclude that for our original question, the factors that affect how fast a finish occurs are, whether the bout is contested in Lightheavyweight and Heavyweight categories (Weightclass), whether the bout ended by KO/TKO or Submission (FinishType), and whether the arm reach difference is large or small (ReachDifference).

Our model has a few limitations. First, the normality assumption is not very well satisfied (as seen in the last provided normal QQ-plot). This definitely affects the estimates and CIs of the coefficients of the predictors we found as significant, meaning that one of the significant predictors might not influence the response in the specified way. Another limitation regarding the intuitiveness and interpretability of our model, is the power transformations. They improve the model, but make the results very hard to understand intuitively. The estimates had very large variances, and about 1/5 of the data were influential points. This means that the coefficient estimates might have given us a wrong image of how each predictor is related to the response in the population level. Lastly, this model was not validated, nor was it compared to other models with different combinations of predictors, thus we have no measure of how good it fits the data, or if there was another model that could yield more accurate results.

# APPENDIX




P1: Scatterplots between each predictor and the response




P2: Influential points(Blue), Leverage points(Red) and Outliers(Green) of each predictor vs the response

## REFERENCES


1) Rainey, C. E. (2009). Determining the prevalence and assessing the severity of injuries in mixed martial arts athletes. North American journal of sports physical therapy : NAJSPT, 4(4), 190–199
2) Telegraph Sport. (2017, August 17). What is UFC, what is MMA and what are the rules? The Telegraph. Retrieved from: https://www.telegraph.co.uk/mma/0/ufc-mma-fighting-rules/.
3) Bounty, P. L., Campbell, B. I., Galvan, E., Cooke, M., & Antonio, J. (2011). Strength and conditioning considerations for Mixed Martial Arts. Strength & Conditioning Journal, 33(1), 56–67. https://doi.org/10.1519/ssc.0b013e3182044304.
4) Miarka, B., Munoz, A. E., Perez, D. I. V., Teixeira, F. G., Brito, C. J. (2020). Ending an MMA combat bout: specific striking techniques which determine the type of outcome. Journal of Martial Arts Anthropology, 20(3), 9- 17.
5) Buse, G. J. (2006). No Holds Barred Sport Fighting: A 10 year review of Mixed Martial Arts Competition. British Journal of Sports Medicine, 40(2), 169–172. https://doi.org/10.1136/bjsm.2005.021295.