Key Determinants To Maximize Total Weight Lifted In Powerlifting*

Ricky Fung

April 19, 2024

This paper examines the characteristics such as age, weight, and equipment usage observed in powerlifting meets from the years 1964 to 2024. A bayesian logistic regression model was created to determine whether the characteristics influence the competitiveness of a lift, specifically whether the lift is over or under 600kg. Results show that age group 24-34, weight class 110-129kg, and equipment usage of single-py, had a greater likelihood of a competitive lift. These findings suggest there are specific characteristics that are optimal for a lifters performance.

1 Introduction

...

2 Data

2.1 Dataset

The data set used in this paper is obtained from OpenPowerlifting (OpenPowerlifting, n.d.). The data is a compilation and standardization of meet data within the sport of powerlifting across an variety of different meets. For example, "World Powerlifting", "Internation Powerlifting Federation", "National Powerlifting Leauge", etc. Organizers of the meets, record their sex, age, weight and lifts performed by each competitor and this data is often made publicly available online.

*Code and data are available at: https://github.com/eatingcorn/powerlift

Table 1: Preview of the cleaned dataset From OpenPowerlifting.

competitive	sex	age_class	weight_class_kg	total_kg
0	Female	24-34	60-69	247.5
0	Female	40-44	70-79	282.5
0	Male	16-17	70-79	267.5
0	Male	35-39	70-79	430.0
0	Male	24-34	70-79	265.0
0	Male	16-17	80-89	275.0
0	Male	55-59	80-89	335.0
0	Female	24-34	100-109	287.5
0	Female	24-34	50-59	212.5
0	Male	24-34	70-79	397.5

The dataset was filtered to contain only individuals that are tested and have not been disqualified to ensure the integrity of the data for the analysis. The majority of attributes in the original dataset were not employed, as the primary focus of this paper lies in analyzing the factors contributing to strength gains in powerlifting. The cleaned dataset contains 8 attributes:

1. `competitive` - Denotes the the competitiveness of the total lifted weight. (0 denotes < 600kg, 1 denotes >= 600kg)
2. `sex` - Denotes the lifter's gender, (Male or Female).
3. `age_class` - Denotes the lifter's age category.
4. `weight_class_kg` - Denotes the lifter's weight category (in kilograms).
5. `total_kg` - Denotes the sum of lifters best bench, squat, and deadlift in kilograms.
6. `equipment` - Denotes the lifter's equipment used in the lift (Wraps, Raw, Unlimited, Single-py, Multi-py).

The attribute `competitive` was introduced into the dataset as a binary variable, indicating the competitiveness of the total lifted weight, and it was incorporated within our logistic model. The attribute weight_class_kg was adjusted to represent a range of weights, aimed at condensing the x-axis values, as the original dataset contained approximately 120 different weight classes.

The `competitive` attribute was chosen as it is the estimated in this paper. The `total_kg` attribute are necessary to visualize the increases and/or decreases in each lift based on the correlated attributes. The attributes `sex`, `age_class` and `weight_class_kg` were chosen as they correlate with an individual's lifting capacity.

The analysis of the data set in (Section 4), will be carried out using the R programming language (R Core Team 2023) and packages; tidyverse (Wickham et al. 2019), arrow (Richardson et al. 2024), janitor (Firke 2023), and dplyr (Wickham et al. 2023), we were able create simulations, clean the dataset and write tests. The figures and tables are generated using ggplot2 (Wickham 2016), knitr (Xie 2014), and png (Urbanek 2022). Package shiny (Chang et al. 2024) was use to create interactive tables (See README).

3 Model

3.1 Model set-up

From Section 4, we observe trends between the average total weight lifted and the attributes sex, age class, weight class, and equipment. We want to further this by investigating whether lifts are considered competitive or not based on the listed attributes. To do this we will create a Bayesian logistic regression model.

$$
\begin{aligned}
y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) && (1)\\
\mu_i &= \beta_0 + \beta_1 \times sex_i + \beta_2 \times age\_class_i + \beta_3 \times weight\_class\_kg_i + \beta_4 \times equipment_i && (2)\\
\beta_0 &\sim \text{Normal}(0, 2.5) && (3)\\
\beta_1 &\sim \text{Normal}(0, 2.5) && (4)\\
\beta_2 &\sim \text{Normal}(0, 2.5) && (5)\\
\beta_3 &\sim \text{Normal}(0, 2.5) && (6)\\
\beta_4 &\sim \text{Normal}(0, 2.5) && (7)\\
\sigma &\sim \text{Exponential}(1) && (8)
\end{aligned}
$$

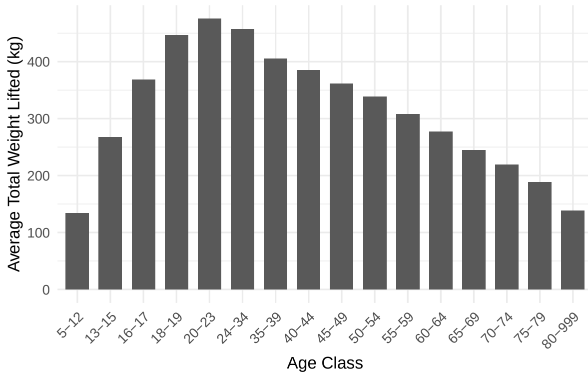
In these equations, y_i represents the competitiveness of the lift (0 not competitive, 1 competitive), π_i represents the probability of the ith observation being a competitive lift, β_0 represents the intercept (log-odds of a competitive lift where the other attributes are 0), β_1 represents the log-odds of a competitive lift depending on the sex, sex_i represents the sex of the ith individual, β_2 represents how much the log-odds of the competitiveness depending on the age class of the lifter, age_class_i represents the age class of the ith lifter, β_3 represents the log-odds of a competitive lift depending on the weight class of the lifter, and equipment_i represents the equipment usage of the ith lifter.

3.1.1 Model justification

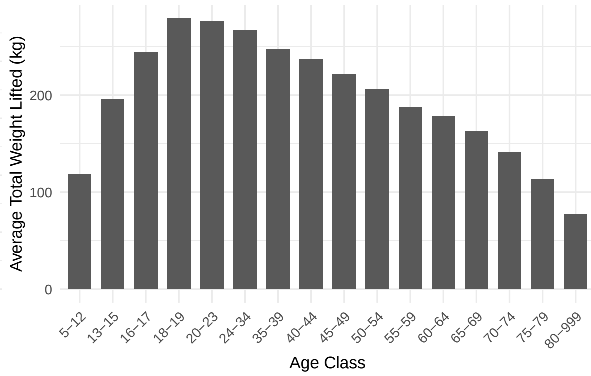
The logistic regression model was chosen in this situation as the outcome of the variable of interest is binary. The logistic regression model also provides easily interpretable results in terms of odds ratios. This is particularly useful in understanding how each predictor variable (such as sex, age class, and weight class) influences the odds of a competitive lift. Other models were consider such as multiple linear regression model but due to the non linearity of some predictors, we decided to proceed with logistic regression. Since logistic does not assume a linear relationship between the predictor variables and the logit of the outcome.

4 Results

4.1 Age Class and Total Weight Lifted (Kg)



(a) Male



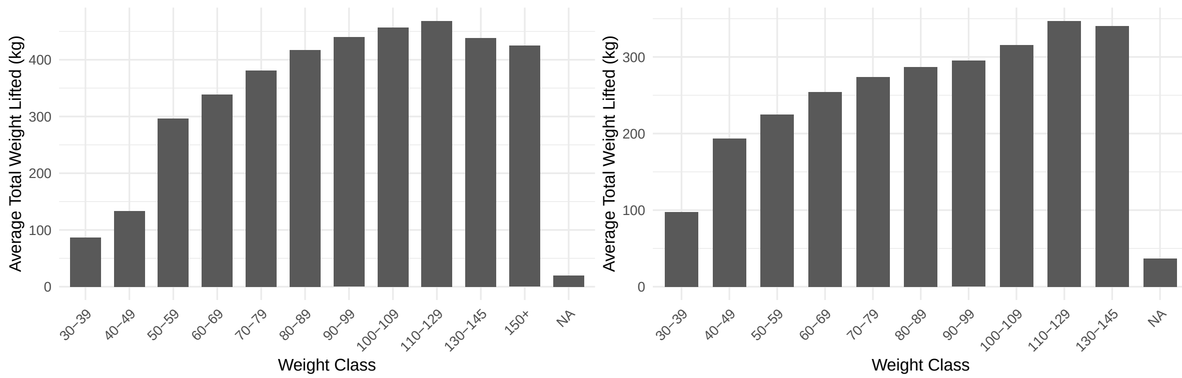
(b) Female

Figure 1: Average Best Bench, Squat, Deadlift based on Age Class

Figure 1 display age classes along the horizontal axis, representing the age ranges during which individuals are lifting weights and the vertical axis shows the average total weight lifted (bench, squat, and deadlift).

For males, the average total weight lifted across age classes demonstrates a steady increase from age class 5-12 to 35-39, peaking notably within the 35-39 age range. Subsequently, there's a gradual decrease observed as age progresses beyond this peak. For females, the average total weight lifted across age classes demonstrates a steady increase from age class 5-12 to 24-34, peaking notably within the 24-34 age range. Subsequently, there's a gradual decrease observed as age progresses beyond this peak.

4.2 Weight Class and Total Weight Lifted (Kg)

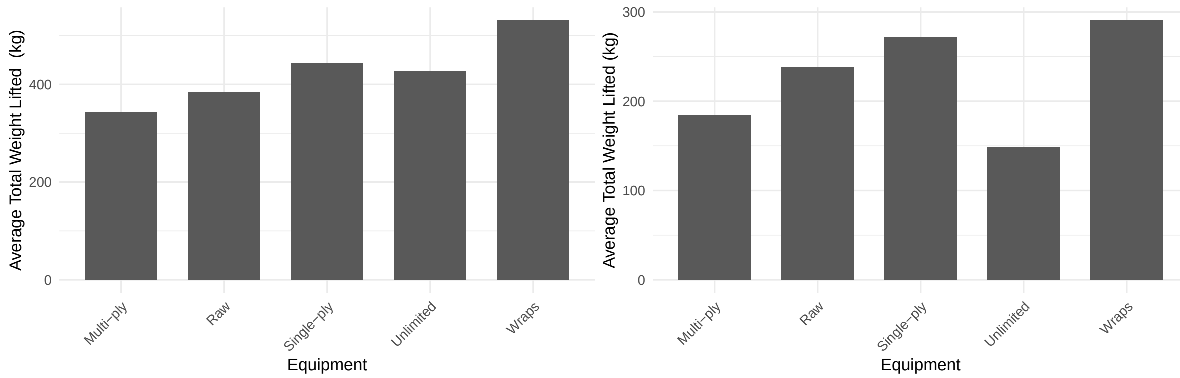


(a) Male (b) Female

Figure 2: Average Best Bench, Squat, Deadlift based on Weight Class

Figure 2 display weight classes along the horizontal axis, representing the weight ranges during which individuals are lifting weights and the vertical axis shows the average total weight lifted (bench, squat, and deadlift). The average total weight lifted across weight classes represented by Figure 2 demonstrates a significant steady increase in the average total weight lifted from across weight classes for both male and female.

4.3 Equipment Used and Total Weight Lifted (Kg)



(a) Male (b) Female

Figure 3: Average Best Bench, Squat, Deadlift based on Weight Class

Figure 3 display Equipment usage along the horizontal axis, representing the equipment usage during which individuals are lifting weights and the vertical axis shows the average total weight lifted (bench, squat, and deadlift). The average total weight lifted across different equipment usage categories, illustrated in Figure 3, indicates a notable increase in average total weight lifted for equipment multi-ply, unlimited, and single-ply, when compared to instances where no equipment is used. Conversely, there is minimal change observed with wraps.

4.4 Model Results

Table 2: The likelihood of a individual to perform a competitive lift

Table 2: A Competitive Lift

Variable	Coefficient	SE
(Intercept)	-7.965	0.642
sexMale	4.286	0.421
age_class16-17	0.673	0.441
age_class18-19	2.152	0.414
age_class20-23	3.245	0.417
age_class24-34	3.056	0.418
age_class35-39	2.514	0.429
age_class40-44	2.152	0.422
age_class45-49	2.094	0.430
age_class5-12	-22.126	21.122
age_class50-54	1.285	0.452
age_class55-59	1.053	0.481
age_class60-64	0.095	0.553
age_class65-69	-1.616	1.219
age_class70-74	-20.635	16.971
age_class75-79	-33.716	30.229
age_class80-999	-41.403	36.204
weight_class_kg110-129	0.039	0.081
weight_class_kg130-145	-0.124	0.214
weight_class_kg30-39	-46.359	44.191
weight_class_kg40-49	-11.014	7.501
weight_class_kg50-59	-3.885	0.376
weight_class_kg60-69	-2.587	0.149
weight_class_kg70-79	-1.789	0.107
weight_class_kg80-89	-0.972	0.079
weight_class_kg90-99	-0.429	0.078
equipmentRaw	0.087	0.209
equipmentSingle-ply	1.229	0.204
equipmentUnlimited	1.204	0.415
equipmentWraps	1.088	0.235
Num.Obs.	15000.000	NA
R2	0.265	NA
Log.Lik.	-4542.366	NA
ELPD	-4567.800	NA
ELPD s.e.	61.100	NA
LOOIC	9135.500	NA

Variable	Coefficient	SE
LOOIC s.e.	122.100	NA
WAIC	9134.900	NA
RMSE	0.310	NA

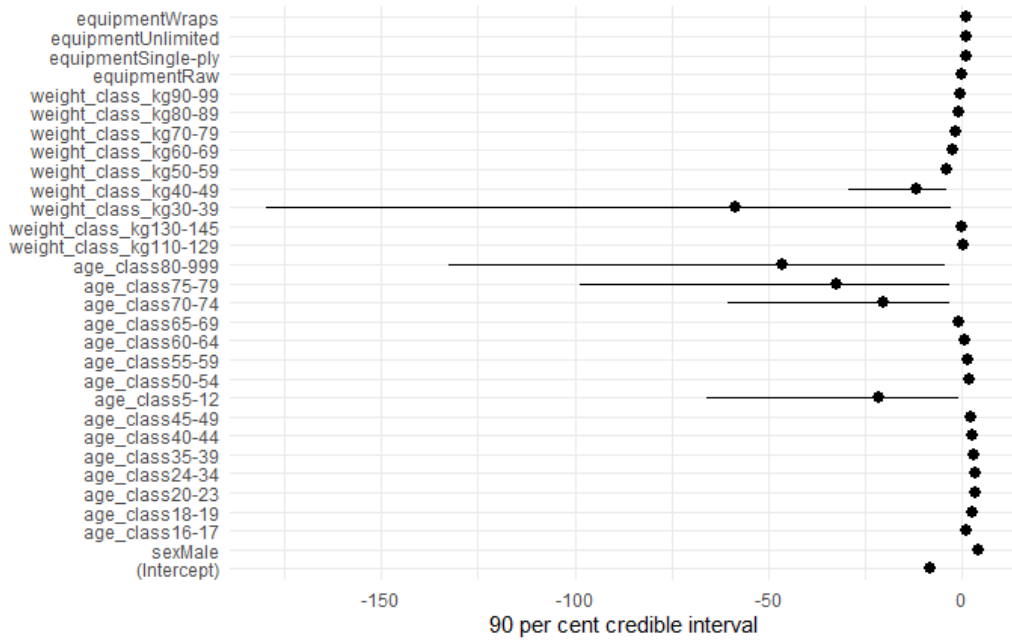


Figure 4: 90% credibility interval for coefficient estimates of predictors of competitive lifts

Table 2 summary results show has an intercept of -7.965, this indicates that baseline performance when all other variables are zero is at -7.965. Being male is associated with a lift increase of 4.286 units, this indicates men are more likely to perform a competitive lift (>= 600kg total weight). Among age groups, the most substantial increase associated with lift is observed in the 20-23 age category, with a coefficient of 3.245 units. This suggests that lifters within this age bracket demonstrate notably enhanced performance. In terms of weight categories, the 110-129kg group displays a modest lift increase of 0.039 units, suggesting a slight advantage in achieving competitive lifts. Different equipment types also influence performance. Notably, specific equipment types show a positive impact on lift performance, with Single-ply and Unlimited gear having coefficients of 1.229 and 1.204 units, respectively.

Figure 4 provides valuable insights into the uncertainty associated with the model estimates. The positive credible interval for the coefficient of the males suggests that being male is likely associated with a statistically significant increase in competitive lifting performance. Age categories such as 50-54, 70-74, 75-79, and 80+ exhibit negative credible intervals, indicating that the estimated performance for these age groups is likely lower than the point estimates provided in the model results. This suggests a level of uncertainty in the effect of age on competitive lifting performance in these age brackets. Similarly, the weight class 30-39 kg and greater displays a negative credible interval, suggesting that the estimated performance for this weight category may be lower than the point estimate provided in the model.

5 Discussion

As evidence by the results, specific characteristics influence the strength of an individual. Specifically age classes, weight classes, and equipment usage effect the strength of an individual.

5.1 Age, Weight, and Equipment Influence The Performance of Lifts.

Figure 1 reveals that men achieve the highest total kilograms lifted (sum of bench, squat, and deadlift) between ages 35 and 39, followed by a steady decrease as the individual ages. Females achieve the highest total kilograms lifted between ages 24-34, followed by a steady decrease as the individual ages. This indicates that the peak performance for males and females of total kilograms lifted occurs during these age ranges. Figure 2 reveals that men and women achieve higher total kilograms lifted (sum of bench, squat, and deadlift) the greater the weight. It suggests that individuals with higher body weights tend to achieve greater total kilograms lifted across the three main lifts. Figure 3 reveals that both men and women achieve an increase in total kilograms lifted, particularly with equipment types such as multi-ply, unlimited, and single-ply. This suggests that the use of equipment not only enhances safety but also boosts strength. Understanding these relationships can inform training strategies tailored to maximize performance, longevity, and individualized optimization.

5.2 Weaknesses

A weakness in this paper, is the reduction of the analysis dataset. The analysis dataset was reduced from 1450643 to 15000 to reduce the run time to create the (**table-x?**), Figure 4. By limiting the analysis to a smaller subset of data, the paper may fail to capture the full variability that may be presented in a larger dataset. Consequently, the conclusions drawn from the reduced dataset may not accurately reflect the estimates and increase the uncertainty surrounding the model coefficients, confidence intervals.

Another weakness in this paper is how surface level it is as paper mainly focuses on basic statistics such as age, weight, and equipment usage. They play significant roles in influencing lift performance, but numerous other factors also can also influence. For example, years of training, years of competing, duration of training sessions, etc. If more advanced statistics were used in this analysis, they could offer deeper insights into the relationships between various factors and lifting performance.

5.3 Next Steps

An appropriate next step in this analysis is to compare the increase in total weight lifted with various scoring systems, such as the Wilks score, GL points, DOTS, and others. Through this comparison, we can identify correlations and discrepancies between the performance metrics used and predict the rankings, thus gaining a deeper understanding of how different scoring systems assess lifting performance.

References

Chang, Winston, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert, and Barbara Borges. 2024. *Shiny: Web Application Framework for r*. https://CRAN.R-project.org/package=shiny.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

OpenPowerlifting. n.d. "OpenPowerlifting." https://openpowerlifting.gitlab.io/opl-csv/bulk-csv.html.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://github.com/apache/arrow/.

Urbanek, Simon. 2022. *Png: Read and Write PNG Images*. https://CRAN.R-project.org/package=png.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.