Seif Emam

Powerlifting data analysis 2020

Abstract: In the powerlifting world, there are numerous factors that can influence the weight lifted by the competitor in the squat, bench, and deadlift. In this research paper, 5 exploratory variables are explored (age, bodyweight, equipment, sex, and testing) and tested to see their effect on the amount of weight lifted by the competitor. The initial hypothesis was that the bodyweight of the competitor would have the biggest effect on the amount of weight they can lift and that the equipment will probably have a major effect. Contrary to my hypothesis, across all lifts, sex had the biggest impact over bodyweight while equipment had a varying effect. Also, in this research, variables were tested to see if they could cluster the data to form weight classes and that did not seem to happen.

Introduction and Background:

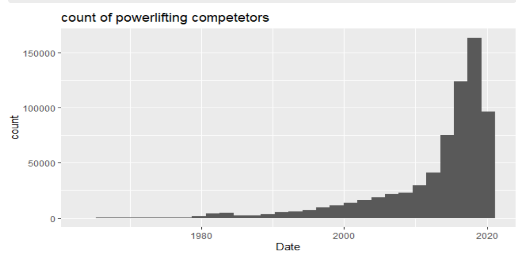
Powerlifting is a very a complex sport in which a lot of physical characteristics as well as external factors can influence the amount of weight lifted by the competitor. Popularity of powerlifting as a competitive sport has exploded over the years (Figure 1). However, what makes a good powerlifter is not yet well understood. There is a lot of room for more in-depth understanding of what influences the competitor’s success.

Figure 1: Number of power lifters over time

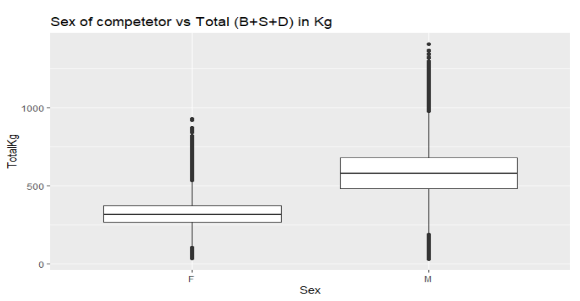
 In this research, the primary aim is to see how much influence the age, weight, sex, the use of equipment, and testing has on the amount of weight lifted by the competitor on each of the squat bench and deadlift.

Figure 2

Methodology:

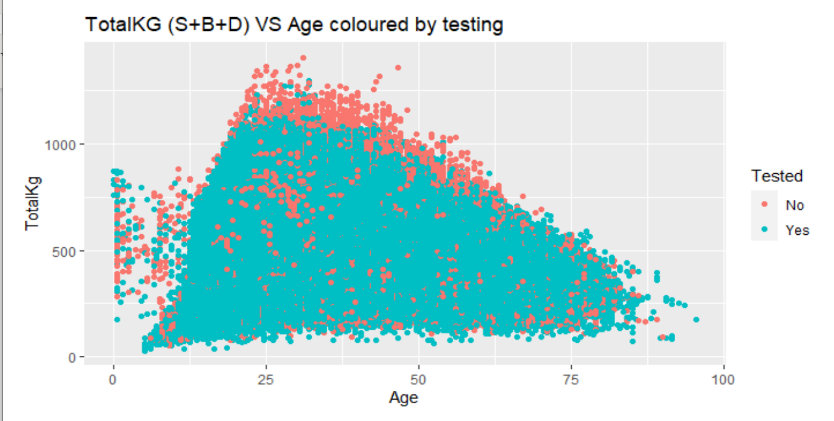
In the initial exploratory analysis conducted, it looked like Sex had major difference on the total amount (squat+bench+deadlift) of weight lifted (fig. 2). The total weight lifted also seemed to peak around age 20-30 (fig. 3). There doesn’t seem to be a very strong relation between the event being tested or not on the amount of weight lifted except for the fact that heaviest weights (1000Kg+) seem to be dominated by non-tested competitors as the red dots indicate.(fig.3)

Figure 3

The data was obtained through a powerlifting database (1). The dataset includes data from 1964 until now. In the exploratory analysis and the visuals shown above(fig1,2,3), all datapoints containing the competitor’s age were included. However, for both the analysis that tests the importance of each variable and the clustering, I decided to analyze the powerlifting meets and data that took place throughout 2020 only. For this analysis, I used cross validation to compare both the Lasso and linear regression model, both using the variables V1(Appendix II) to predict the weight lifted on each of the lifts. Both models had very similar RMSE’s and R squared scores on each of the squat, bench, and deadlift, so I decided to go with the Lasso model on all 3 of them just for consistency. Multi-ply and single-ply suits as well as the unlimited factor were grouped together under the name suits\_more in the equipment’s column (Appendix I).

# Clustering and Weight classes:

The 2020 data set was divided into 2 data frames based on sex. Pam clustering was used to cluster the data in each of these data sets into 3 clusters based on variables V2 (Appendix III). A 2-way table was then set up to compare the frequency of each weight class in each cluster.

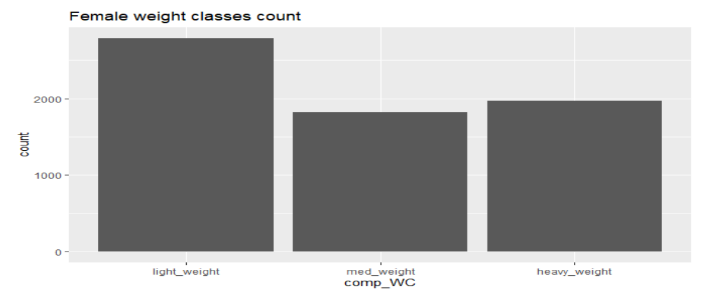
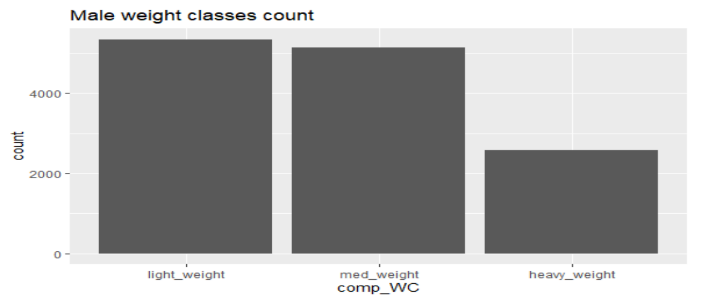
**** The original weight class column included over 34 factors since different federations use different weight classes. Weight classes were reduced to three: below 83kg, between 83kg and 105kg, and 105kg+ for men, and below 63kg, between 63kg and 72kg, and 72kg+ for women labeled as ‘light’, ‘med’ and ‘heavy’ weight respectively. I arrived at these weight class selections using the weight classes list used by the IPF (2)where I chose 3 weight classes from them that would split the data in 3 portions that are comparable in size as the histograms show (fig. 4,5).

Figure 5(2020 only)

Figure 4 (2020 only)

Results:

Figure 7

# Squats:

Figure 6

As the plot shows (Fig 7), Sex is by far the most important factor influencing the squat weight lifted, Bodyweight is a far second, followed closely by whether the competitor is wearing a suit or not. All else being constant, if the meet is drug tested, that seems to reduce the amount of load lifted by 14 units. (Fig 6)

Figure 9

# Bench:

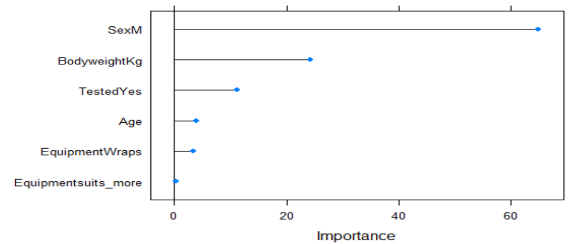
Figure 8

With the bench, wearing suits or using equipment beyond knee wraps seems to have a very important effect just behind the sex. (Fig.9)

# Deadlifts:

Figure 10

Figure 11

With deadlifts, the equipment matters much less than the squat or the bench (Fig.11) and being older seems to be more deleterious to the deadlift than to the squat or the bench as the -3.83 coefficient shows. (Fig.10)

In general, we can see clearly that being a male and having a higher bodyweight are very crucial in increasing the weight lifted across all three lifts with deadlifts being most affected by sex ( increase of 64 units in the weight lifted when being a male) while squats being most affected by bodyweight ( increase of 28 units in weight squatted when increasing the bodyweight by 1 unit).

Wearing a ply suit seems to cause a great increase in the weight squatted and benched while having very little effect on the deadlift.

Clustering and weight classes:

|  |  |  |  |
| --- | --- | --- | --- |
| Male cluster frequency (table 1) | Lightweight | Midweight | Heavyweight |
| 1st Cluster | 397 | 1501 | 1433 |
| 2nd cluster | 1991 | 2566 | 890 |
| 3rd cluster | 2962 | 1075 | 251 |

Male clusters (table 1): There does not seem to be any significant trends across all clusters, but the 3rd cluster seems to be light weight dominant while the 2nd cluster seems to be med weight dominant.

|  |  |  |  |
| --- | --- | --- | --- |
| Female cluster frequency (table 2) | Lightweight | Midweight | Heavyweight |
| 1st Cluster | 358 | 479 | 810 |
| 2nd cluster | 1396 | 501 | 381 |
| 3rd cluster | 1034 | 843 | 779 |

# Female clusters (table 2):

Just like their male counterparts, the female clusters did not show extremely strong trends, but the first cluster seems to be clearly heavyweight dominant, while the 2nd cluster seems to be light weight dominant.

Discussion/Assumptions/Limitations:

It is important to note that powerlifting meet labeled as not tested does not mean that all competitors are on drugs and a meet labeled as tested does not guarantee that all the participants are tested. The analysis was done on the meets done this year, and due to Covid situation that makes this year a unique case, we should be careful when trying to extrapolate results from this analysis to data about meets in previous years.

# Variables importance comparison:

The results seem to indicate that being a male is the more important than all of the other factors considered in lifting more weight. We can just speculate that this might be a result of the hormonal composition that males seem to have, but this just a speculation. More research is needed in this area, to give us a clear picture of why sex might be this important in influencing the ability to lift heavier loads.

# Clustering and weight classes:

There are no significant trends in the male or female clusters to show that weight classes would “naturally” take place. This might seem as a result that undermines the concept of weight classes; however, the results from a method like clustering using only 9 variables used on a dataset of modest size ( around 13k rows for men and 7k for women) are not conclusive by any means. More research is needed to give us a clearer picture about the importance or lack of weight classes in powerlifting.

Appendix:

I) Original equipment options in the dataset:

- Raw: Bare knees or knee sleeves.

- Wraps: Knee wraps were allowed.

- Single-ply: Equipped, single-ply suits.

- Multi-ply: Equipped, multi-ply suits (includes Double-ply).

- Unlimited: Equipped, multi-ply suits or rubberized gear (like Bench Daddies).

- Straps: Allowed straps on the deadlift (used mostly for exhibitions, not real meets).

II) variables V1: used in variables importance analysis

(age, bodyweight, equipment, sex, and testing)

III) Variables V2 : used in clustering and weight classes:

(age, bodyweight, equipment, sex, testing, bestbenchKg, bestsquatKg,bestdeadliftKg)

# References

(1). Retrieved from https://www.openpowerlifting.org/data

(2). Retrieved from http://www.usaplmn.com/wp-content/uploads/2015/07/IPF-USAPL-Weight-classes.pdf