What are Your Chances of Squatting Three Plates?

Luis Sebastian Weekes,

Kenyon College,

Stat 206

May 11, 2023

Introduction:

In the exercising world there are many benchmarks. One popular benchmark is "squatting three plates". Squatting three plates is squatting 315 pounds. In this study, I am interested in variables that determine the chances of you being able to squat three plates. The five variables that I will be looking at are: **age, gender, height, weight, how long a person has been training for and a person's workout frequency**. I will also be checking to see if there are any logical interactions. The age variable could be very interesting, because we have a wide range of ages in our data set—and we know that people get stronger as they get older, but eventually start getting weaker past a certain age. According to Hurley (1995), people reach maximum strength between their second and third decade and maintain a significant amount of their strength in to their fourth or even fifth decade, with most decline starting in their fifth decade for sure. Gender will be an important factor to consider since males are generally stronger than females across the board. For height, I hypothesize that shorter people in general might be able to squat more than taller people, because of the fact that they only need to move the weight for a shorter distance. Furthermore, Heeran (2022), argues that taller people have a hard time squatting because of their higher center of gravity which makes it more difficult to control heavier weights. For bodyweight, I hypothesize that heavier people will be found to squat over 315 more often, because they have more muscle mass.

The data that I will be using was collected by Sam Swift by web scraping from official Cross fit Databases, Swift (2015). The data was collected in 2015 "to bring more data to the cross fit open," (Swift 2015). For the purpose of this paper, the population of people who can back squat 315 pounds and above will be called 315ers, "how long" will be how many years a person has been doing cross-fit for and schedule will be a measure of workout frequency.

Exploratory Data Analysis:

For each of the quantitative predictors, I generated a box plot, and ran a t-test in order to determine the validity of considering them in the model.

Age EDA:

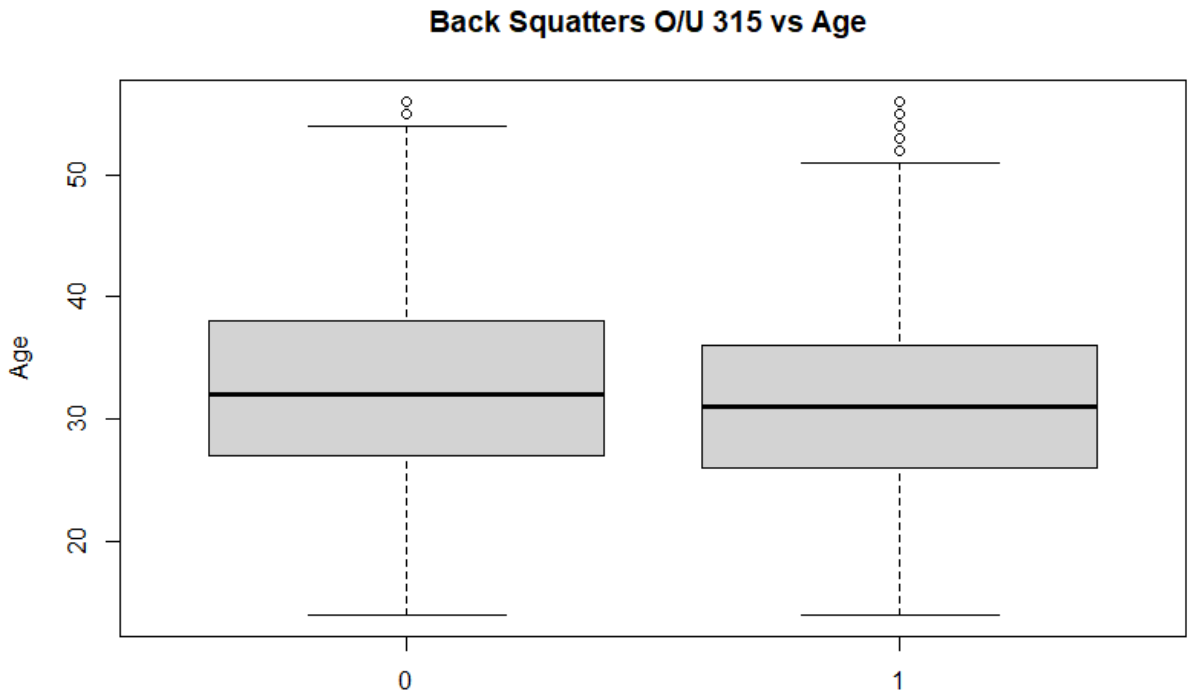


Figure 1

Given Figure 1, it seems as though the population of people who back squat 315 lbs and over might have a slightly smaller mean age than non-315ers. Upon running a t-test, with t = 22.888, df = 54584, p-value < 2.2e-16, we can confirm that that two populations are statistically different. Hence, adding an age predictor to our model might be beneficial.

Weight EDA:

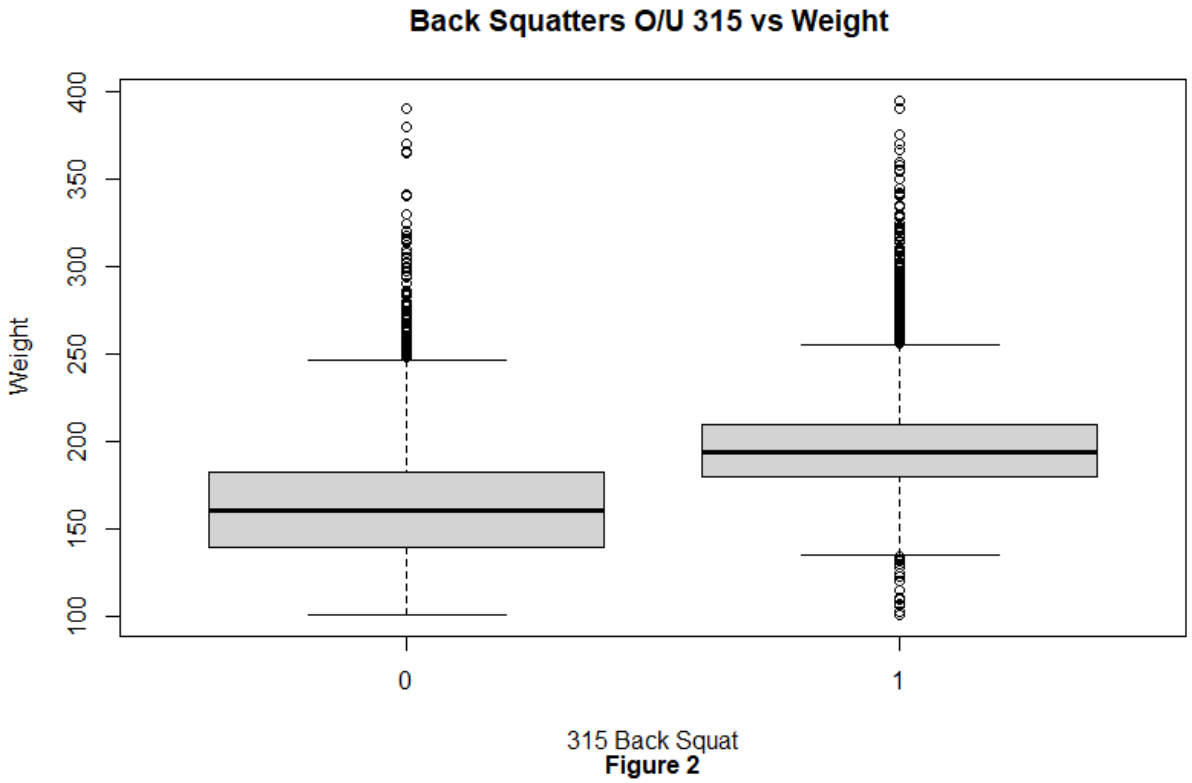


Figure 2

Upon looking at figure 2, we can see that there is a stark contrast between 315ers and non-315ers in terms of weight. The mean weight of 315ers is much higher than non-315ers. We can confirm that the mean weight between 315ers and non-315ers is statistically different with t = -143.05, df = 55160, p-value < 2.2e-16. This seems to follow our initial prediction.

Height EDA:

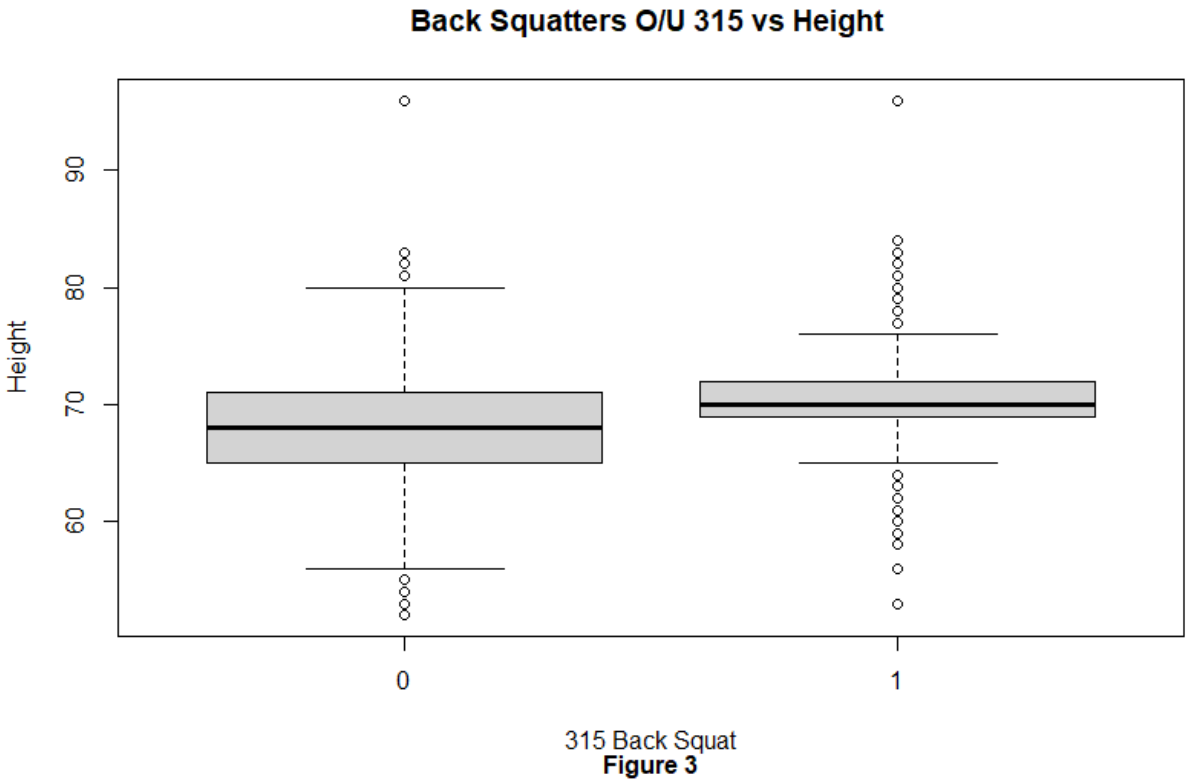


Figure 3

The box plot of height vs 315ers and non-315ers shows that the mean height of 315ers is larger than non-315ers; this difference in means is also statistically significant with t = -98.102, df = 58518, p-value < 2.2e-16 upon conducting a t-test. This runs contrary to my initial prediction, but this may change once we account for other factors.

Gender EDA:

Gender Proportion Table			
back_315	--	Female	Male
0	0.67	0.99	0.44
1	0.33	0.01	0.56
Figure 4			

As you can see from figure 4, when comparing Males to Females, the proportion of Males that can back squat 315 pound is much larger than females.

How Long EDA:

How long Proportion Table						
back_315	1-2 years	2-4 years	4+ years	6-12 months	Decline to answer	Less than 6 months
0	0.61	0.56	0.44	0.67	0.45	0.68
1	0.39	0.44	0.56	0.33	0.55	0.32
Figure 5						

According to figure 5, the proportion of people who can squat 315 pounds increase as the years that you have been training for increases.

Training Frequency EDA:

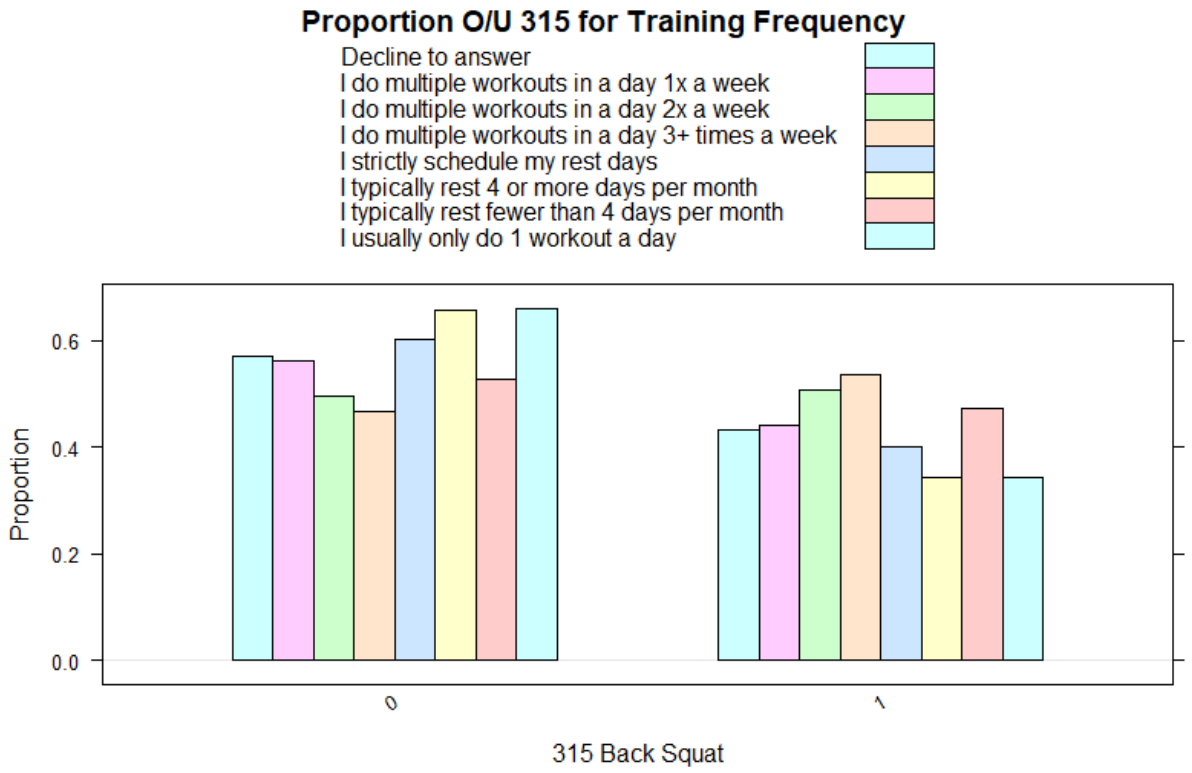


Figure 6

Upon reviewing Figure 6, as the frequency of workouts increases, the proportion of the people who can back squat over 315 pounds for the respective categories increases.

Interactions:

After examining Interaction box plots and empirical logit plots, no apparent interactions surface.

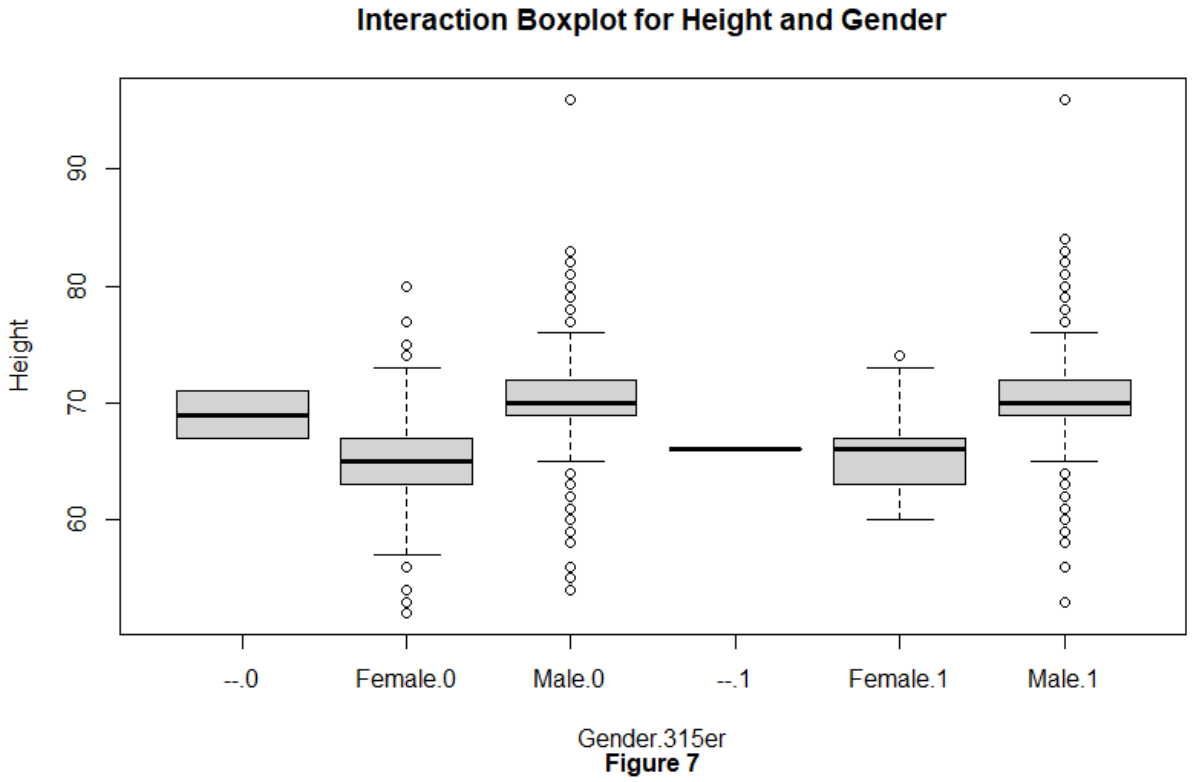


Figure 7

All of the interaction box plots for the Gender predictor look like a version of figure 7. There is no clear tell of any interaction occuring.

Figure 8: Logit Plot for Weight and How long

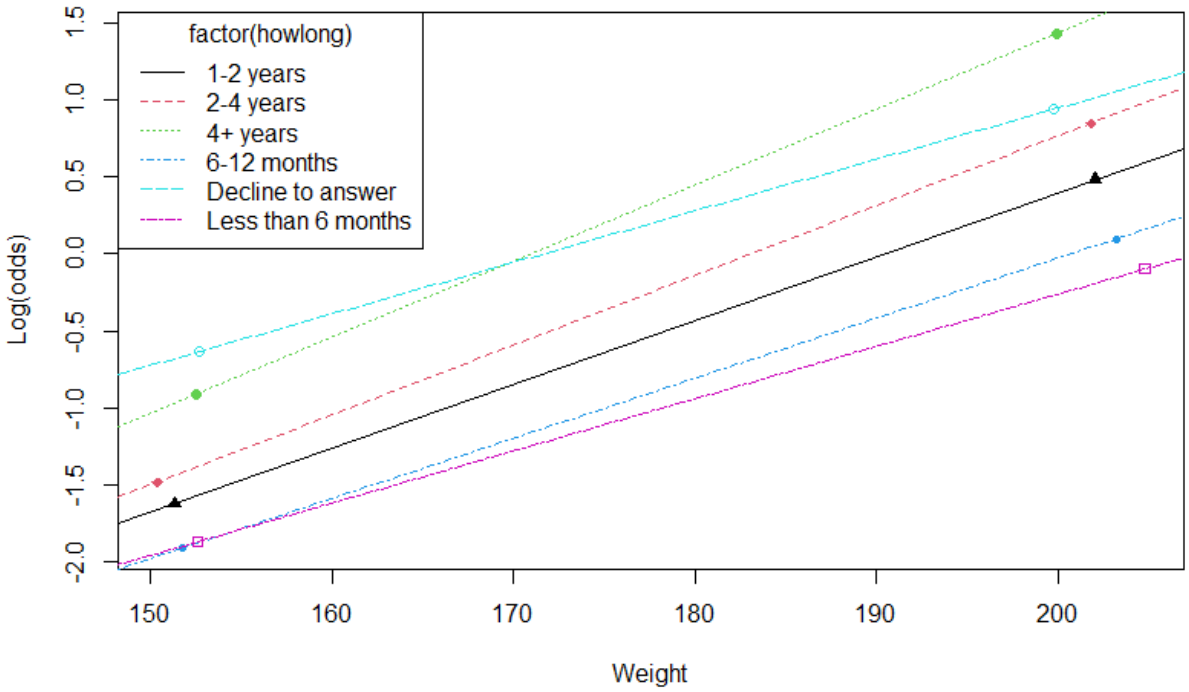


Figure 8 is an example of how the interactions between Weight and the categorical variables looked. As you can see, all of the categories follow the same trend, therefore there is no clear interaction occurring. Furthermore, the rest of the empirical logit plots look similar to figure 8. Empirical logit plots were generated for weight, age and height and their interactions with schedule and how long.

Transformation Checks and Conditions

Figure 9: Logit Plot for Age

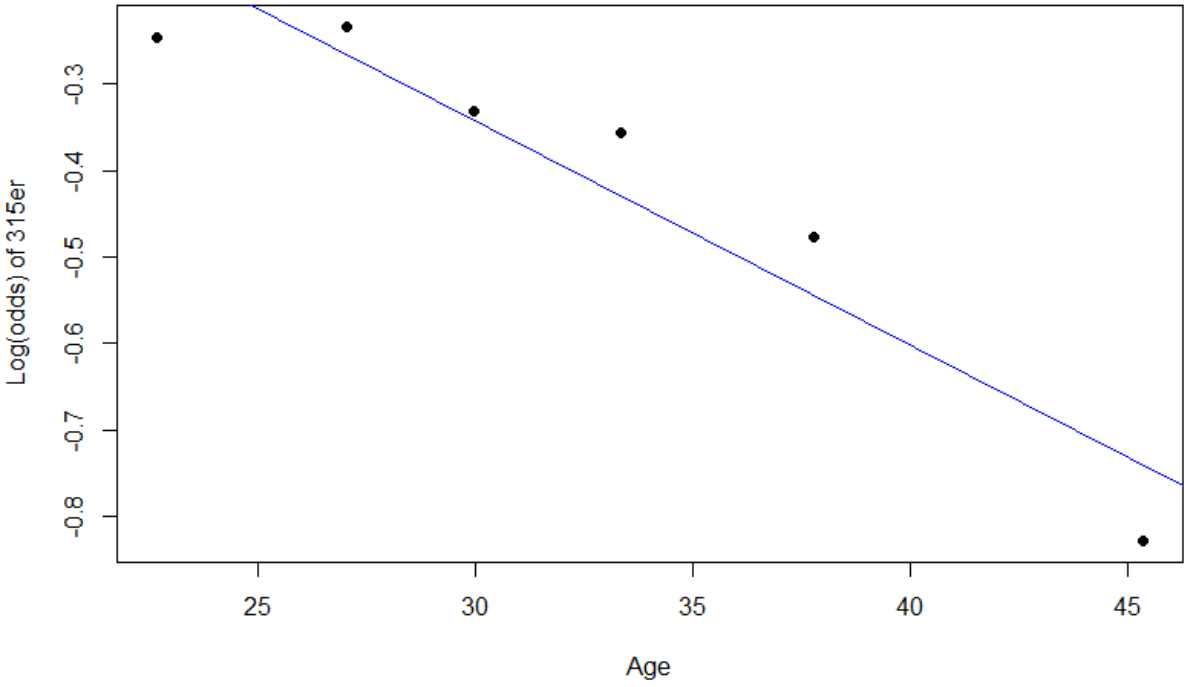


Figure 10: Logit Plot for Weight

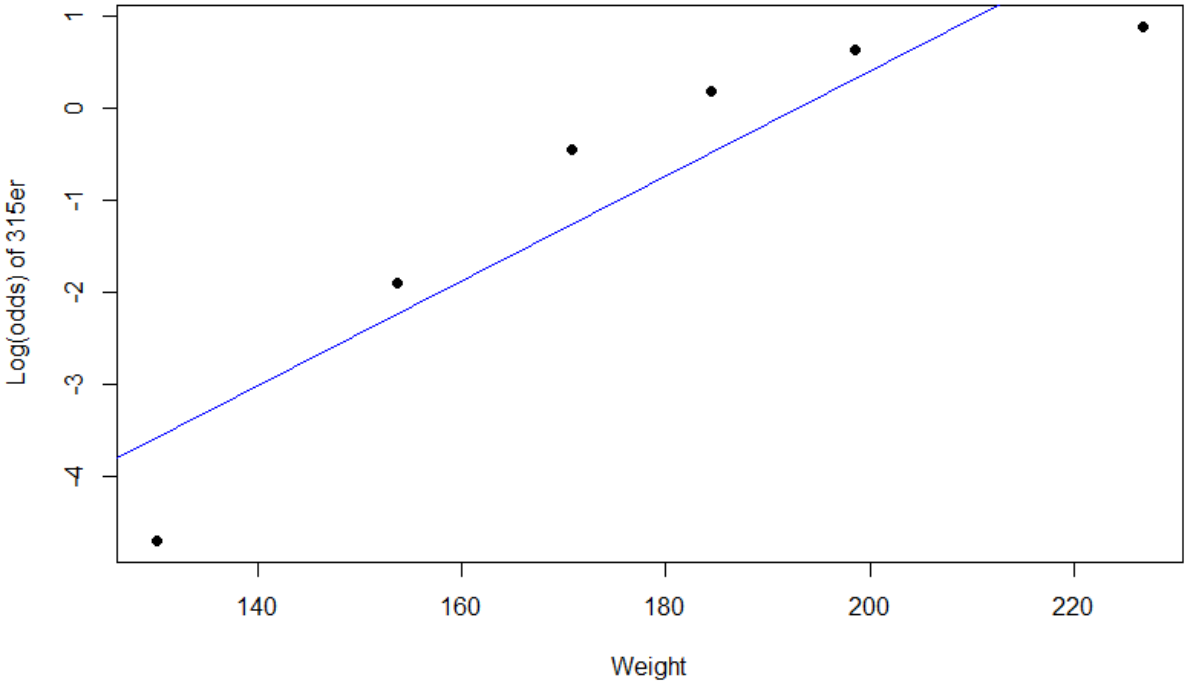
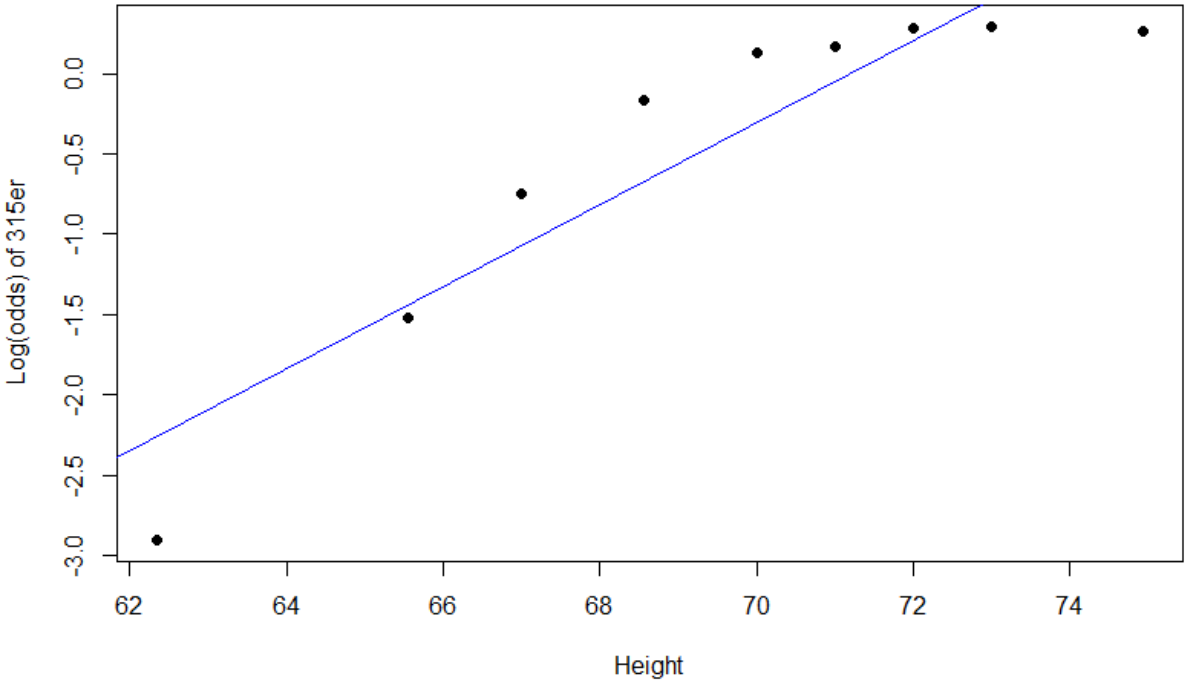


Figure 11: Logit Plot for Height



Figures 9,10 and 11 indicate that no transformations are needed and that the linearity conditions are met in order to proceed with the model. The condition of randomness becomes somewhat questionable if we were to try and apply this data to the general public, because the data is specifically taken from cross-fitters. That being said many people do cross-fit and cross-fit is not too different from other forms of exercise. Furthermore, it is not unreasonable to assume that many different types of people take part in cross fit. We shall therefore proceed cautiously while assuming that this model passes this randomness criterion. In terms of independence, this model will pass, because a person's ability to squat is solely dependent on their genes and lifestyle and has nothing to do with the decisions or genes of other people.

Model Selection:

In order to select a model, I started off with the predictors of age, height, gender, weight, how long and schedule for my full model. I took advantage of the backwards stepwise function that minimizes AIC while removing predictors. Unsurprisingly given earlier EDA, each predictor was significant for the model.

Final Model:

$$P(315er) = \frac{e^{\beta 0+\beta age*age + \beta height*height +\beta gender*gender + \beta weight*weight + \beta howlong*howlong + \beta schedule*schedule}}{1 + e^{\beta 0+\beta age*age + \beta height*height +\beta gender*gender + \beta weight*weight + \beta howlong*howlong + \beta schedule*schedule}}$$

The final model was checked for multi-collinearity (figure 12).

Figure 12: Multi-Collinearity Check

Term	VIF	VIF 95% CI	Increased SE	Tolerance	Tolerance 95% CI
age	1.18	[1.17, 1.19]	1.09	0.85	[0.84,0.86]
height	1.61	[1.59, 1.63]	1.27	0.62	[0.62, 0.63]
gender	1.05	[1.05, 1.06]	1.03	0.95	[0.94, 0.96]
weight	1.67	[1.65, 1.69]	1.29	0.6	[0.59, 0.61]
howlong	1.23	[1.22, 1.24]	1.11	0.81	[0.80, 0.82]
schedule	1.16	[1.15, 1.17]	1.08	0.86	[0.86, 0.87]

Results:

Figure 13 is a display of the resulting coefficients for the model.

Figure 13: Coefficient for model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	6.431099	1.347524	4.773	1.82E-06
age	-0.07084	0.001602	-44.22	< 2e-16
height	-0.16742	0.004833	-34.641	< 2e-16
genderFemale	-5.69029	1.314608	-4.329	1.50E-05
genderMale	-0.68319	1.310597	-0.521	0.602169
weight	0.044736	0.00064	69.888	< 2e-16
howlong2-4 years	0.523345	0.028159	18.585	< 2e-16
howlong4+ years	1.235926	0.042979	28.757	< 2e-16
howlong6-12 months	-0.54229	0.032216	-16.833	< 2e-16
howlongDecline to answer	0.740026	0.211184	3.504	0.000458
howlongLess than 6 months	-0.78158	0.04057	-19.265	< 2e-16
scheduleI do multiple workouts in a day 1x a week	0.346298	0.130712	2.649	0.008065
scheduleI do multiple workouts in a day 2x a week	0.594216	0.126374	4.702	2.58E-06
scheduleI do multiple workouts in a day 3+ times a week	0.578963	0.123642	4.683	2.83E-06

Experience Vs Limited Experience

The odds of back squatting 315 or more is 7.5 times higher for a person who has been doing cross-fit for four years or more compared to a person who has been doing it for less than six months (all else equal)

Often Vs Infrequent:

The odds of back squatting 315 or more is 1.26 times higher for a person who works out 3+ times a week compared to a person who only works out once a week (all else equal).

Gender:

The odds of back squatting 315 or more is 295.97 times higher for a male than a female (all else equal).

6 ft vs 6ft 8 inches:

The odds of back squatting 315 or more is 3.81 times higher for a person who is 72 inches in height compared to someone who is 80 inches in height (all else equal).

220 vs 180:

The odds of back squatting 315 or more is 5.99 times higher for a person who is 220 pounds compared to someone who is 180 pounds (all else equal).

Pearls before Swine or Age before Beauty?

The odds of back squatting 315 or more is 9.65 times higher for a 23 year old compared to 55 year old (all else equal).

How Well Does the Model Do?

Figure 14: Testing the model

	actual non-315er	actual 315er	
prediction	FALSE	TRUE	Totals
TRUE	666	1811	2477
FALSE	2856	535	3391
Totals	3522	2346	

According to figure 14, the model predicted 1811 out of the 2346 observations that were true while predicting an extra 666 observations to be true while they were false. Also, the model predicted 2856 out of the 3522 observations that were false while predicting an extra 535 observations to be false while they were true. Therefore, this model does seem to have a decent amount of predictive power, but more accuracy is desired.

Discussion:

First of all, the odds ratios examined in the results must be taken with a pinch of salt, because the odds ratios remain constant throughout all values of X. For instance, the comparison between a 13-year-old and a 45-year-old would have the same interpretation as a 23- and 45-year-old which is not logical. In addition, the model lacks some predictive power. One major reason that the model lacks predictive power is because people are so unique. Even though we accounted for many variables-- we failed to account for several other genetic variables such as dominant muscle fiber type, and we failed to account for lifestyle choices such as diet. As mentioned earlier, we may also have issues using this model to represent a further population since this data was taken from cross fitters specifically and not the general public. That being said, the model has some predictive value as seen in figure 14.

Conclusions:

This paper took data from a cross fit database and fit a logistic regression for the probability of lifting over 315 pounds with **age, gender, height, weight, how long a person has been training for and a person's workout frequency** as predictors. According to our model's coefficients, as you increase in age, your probability of squatting three plates decreases, males are far more likely to lift three plates, shorter people are more likely to lift three plates, heavier people are more likely to life three plates, people who have been working out longer are more likely to lift three plates and people who work out more often are more likely to lift three plates (all else equal).

References:

Hurley, Ben F. "Age, gender, and muscular strength." *Journals of Gerontology-Biological Sciences and Medical Sciences* 50 (1995): 41-44.

Heeran, Kris "Why Squatting Is Harder for Tall People." GymPosts, 8 Aug. 2020, gymposts.com/training/why-squatting-is-harder-for-tall-people/.

Swift, Sam "CrossFit Games Data 2012-2015" Sam Swift, 18 Mar. 2015, https://swift.pw/crossfit-games-data-2012-2015/#mission-part-1-bring-crossfit-data-mainstream--success