An insight into strength dependency on weight via regression analysis by Ehsan Ghani

I will first show my entire R-code and you can find the analysis below.

```
> library(readxl)
> BenchAverageData <- read_excel("~/Downloads/Dataset for bench press strength
standards for Adult men by Ehsan Ghani.xlsx",
+     range = "A17:B28")
> View(BenchAverageData)
> attach(BenchAverageData)
> plot(`Body Weight (kg)`, `Bench press standards per class of body weight (kg) -
Average`, main = "Scatterplot for average")
> cor(`Body Weight (kg)`, `Bench press standards per class of body weight (kg) -
Average`)
[1] 0.9417726
> Regression.model <- lm(`Bench press standards per class of body weight (kg) -
Average` ~ `Body Weight (kg)`)
> summary(Regression.model)

Call:
lm(formula = `Bench press standards per class of body weight (kg) - Average` ~
    `Body Weight (kg)`)

Residuals:
    Min      1Q  Median      3Q     Max
-10.531  -4.989   1.869   5.267   7.950

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        42.62755    6.86517   6.209 0.000157 ***
`Body Weight (kg)`  0.62692    0.07461   8.402 1.49e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.077 on 9 degrees of freedom
Multiple R-squared:  0.8869,  Adjusted R-squared:  0.8744
F-statistic:  70.6 on 1 and 9 DF,  p-value: 1.492e-05

> abline(Regression.model, col=4, lwd=3)
> anova(Regression.model)
Analysis of Variance Table

Response: Bench press standards per class of body weight (kg) - Average
                   Df Sum Sq Mean Sq F value    Pr(>F)
`Body Weight (kg)`  1 3536.0  3536.0  70.601 1.492e-05 ***
Residuals           9  450.8    50.1
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> stdres<- rstandard(Regression.model)
> print(stdres)
         1          2          3          4          5          6
-1.4864868 -1.0645996 -0.5043759  0.2843252  0.6513863  1.1083291
         7          8          9         10         11
 1.1786018  0.9248828  0.5985440 -0.3245869 -2.0230776
> plot(`Body Weight (kg)`, stdres,  main="Std residuals versus explanatory variable")
> fits<- fitted(Regression.model)
> plot(fits, stdres, main="Std residuals versus fits")
> qqnorm(stdres, main="Q-Q Plot")
> qqline(stdres)
> shapiro.test(stdres)

      Shapiro-Wilk normality test

data:  stdres
W = 0.91857, p-value = 0.3069
```

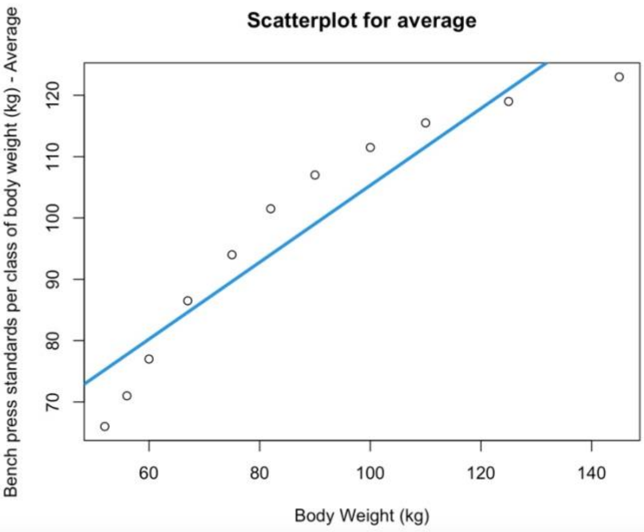
When writing the code above, I had assigned 'Body Weight (kg)' to be the Explanatory variable, and 'Bench press standards per class of body weight (kg) – Average' to be my Response variable. As I had correctly imported my dataset from Excel, assigning variables was as simple as correctly plotting for what necessary.

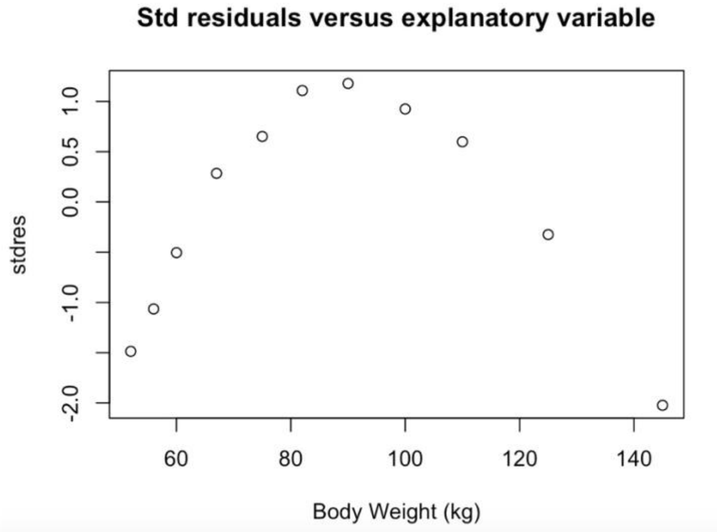
The dataset focuses on the average bench press standards (kg) for Adult men across various weight classes. Whilst the assumption that the more you weigh, the more you could potentially lift, is trivial in nature; I wanted to test just how far we can consider something to be trivial when it comes to human characteristics, like strength's dependency on body weight.

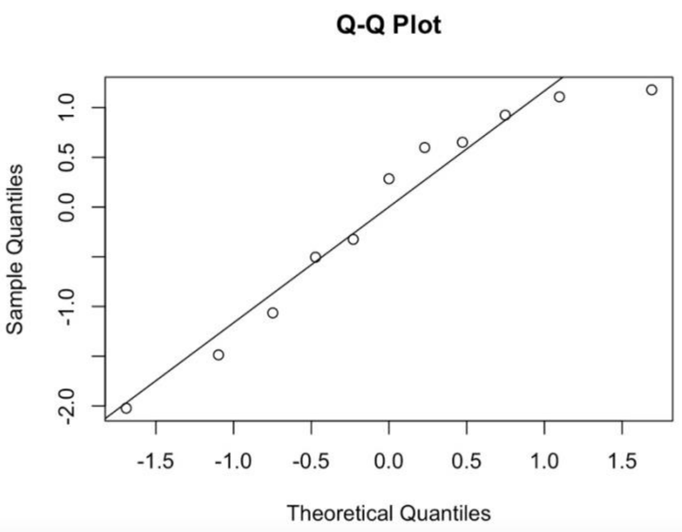
```
> summary(Regression.model)

Call:
lm(formula = `Bench press standards per class of body weight (kg) -
Average` ~
    `Body Weight (kg)`)

Residuals:
    Min      1Q  Median      3Q     Max
-10.531  -4.989   1.869   5.267   7.950

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        42.62755    6.86517   6.209 0.000157 ***
`Body Weight (kg)`  0.62692    0.07461   8.402 1.49e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.077 on 9 degrees of freedom
Multiple R-squared:  0.8869,	Adjusted R-squared:  0.8744
F-statistic:  70.6 on 1 and 9 DF,  p-value: 1.492e-05
```

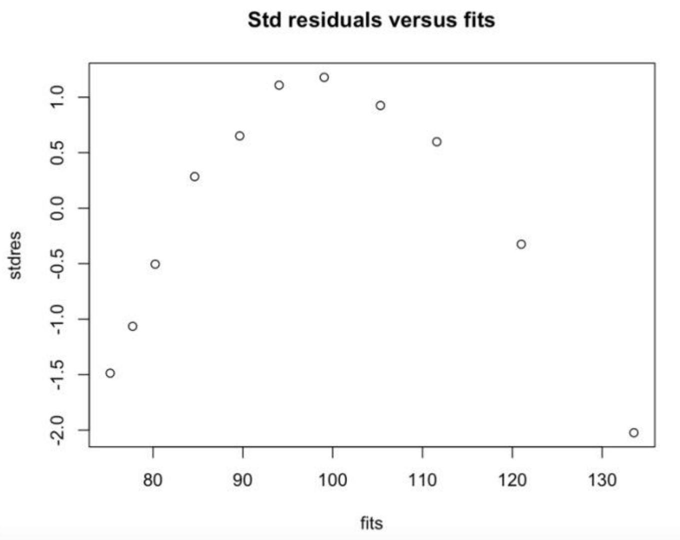
The two regression parameters:
The value for my intercept is highlighted in the blue box and my slope is highlighted in the red box. From this, I have the equation:
y = 0.62692x + 42.62755

It is evident that the intercept and slope are both positive. Even though the slope parameter appears to be very near zero. One interpretation could be that for every 1kg increase in weight, the model suggests an increase in bench press standards per class of body weight by 0.62692 times.










```
> plot(`Body Weight (kg)`, `Bench press standards per class of body weight (kg)
- Average`, main = "Scatterplot for average")
> stdres<- rstandard(Regression.model)
> print(stdres)
         1          2          3          4          5          6 
-1.4864868 -1.0645996 -0.5043759  0.2843252  0.6513863  1.1083291 
         7          8          9         10         11 
 1.1786018  0.9248828  0.5985440 -0.3245869 -2.0230776 
> plot(`Body Weight (kg)`, stdres,  main="Std residuals versus explanatory
variable")
> fits<- fitted(Regression.model)
> plot(fits, stdres, main="Std residuals versus fits")
> qqnorm(stdres, main="Q-Q Plot")
> qqline(stdres)
> shapiro.test(stdres)
```

Onto the evidence from data gathered, there are a few points of interest. When we examine the original scatter plot, we can see that it displays a slight curve rather than a strictly linear relationship between bench press standards per weight class and body weight. In the context of strength and body weight, a non-linear relationship might be possible. As body weight increases, strength might initially increase proportionally, but eventually reach a constant or even decline due to other factors like limitations in muscle mass.
From the residual plots, there is clear evidence of non-linearity, from the symmetric trend we can confirm that there is an outright nonlinear

relationship between the variables, suggesting that the explanatory variable systematically affects the response variable.

Normality of the dataset and topic:

Shapiro-Wilk normality test.

data: stdres
W = 0.91857, p-value = 0.3069

It is evident that the test fails to reject the null hypothesis of normality, as it is greater than 0.05. Furthermore, the Q-Q-plot above doesn't show any strong deviations from normality.

Strong evidence of normality and a lack of linearity is the perfect explanation for human characteristics like weight and strength. As in theory, we are seemingly linear in what we can achieve i.e. the more you train, the more you can lift. But when practicality is implied, genetics defining potential ability etc, it is apparent that where there is life; there are limitations. In retrospect the dataset could only follow a normal distribution. There's no practicality towards, for example, a uniform distribution, as human ability is neither infinitely measurable nor equally likely.

The last point of interest to conclude was from my initial summary of the regression model:

Multiple R-squared: 0.8869, Adjusted R-squared: 0.8744

The R-squared is a goodness of fit measure that indicates the variability in the observations is well explained by the fitted model. To which an R-squared value of 0.8869 or 88.69% indicates that body weight explains approximately 88.69% of the variance of average bench press standards per class of body weight. During the model process, this result had been initially misleading as the underlying relationship between variables was becoming more and more non-linear. Perhaps because the relationship is non-linear, then using quadratic linear regression or other transformations to capture the non-linearity might have been optimal.