Statistical Report

Phan Vinh Phu
StudentID : 45747989

Abstract

Olympic weightlifting requires technical skills, explosive power, strength, and coordination. Weightlifters can be competitive within a range of morphological characteristics due to competition body weight classes. To date, no studies have examined when sex and age differences arise in weightlifting and the impact of body weight on performances at different ages.

Introduction

Women athletes are known to be less strong and powerful than equally trained men. The main factor accounting for gender differences in maximal strength,indeed, has been identified as age. To the best of our knowledge, no studies to date have compared age of male and female athletes or the relation between body weight and the maximum weight athletes can lift. Therefore, the first aim of this study is to answer whether there is any difference in the average age of female and male weightlifters. Beside that, the second aim of the present study was to assess the relationships between the body weight of weightlifters and the weight lifted in sports with high expressions of strength and power.

Methods

A random sample of 261 weightlifters ($age = 28.5 \pm 5 years$; $bodyweight = 97.54 \pm 14.7 kg$; $weightlifted = 146.7 \pm 30.8 kg$) have been collected. The sample includes information about weightlifters' gender, body weight, age, and the maximum weight they have lifted. The first research question is "Is there any difference in the average age of female and male weightlifters?". To answer this question, I apply the two independent sample t-test method. The test is only valid if we can reasonably assume that both samples are drawn from normal distributions. Therefore, the data is split by gender into two small samples. Then, the comparative box plot and the histogram are drawn to check the distribution of the age column filtered by gender.The participants in the present study were 129 women(Female: $age = 28.54 \pm 5.12 years$) and 132 men (Female: $age = 28.44 \pm 4.86 years$).

The second research question is about the relationships between the body weight of weightlifters and the weight lifted in sports. For the second question, a scatter plot is used to determine the linear relation between two variables (bodyweight and Weightlifted). Secondly, I have chosen to use the hypothesis test for the linear relationship to test the relation between them. In this study, the sample will be split into two groups separated by gender.

Result 1

The following tables show information about variables in the sample dataset and descriptive statistics to give an overview.

Table 1: Variable Description

Variable	Description
Id	Subject ID
Gender	Either "female" or "male"
BodyWeight	The weight of the subject
Age	The age of the subject in years
Weightlifted	The maximum weight lifted by the subject in an unsepcified exercise

Table 2: The First Five Rows of Sample

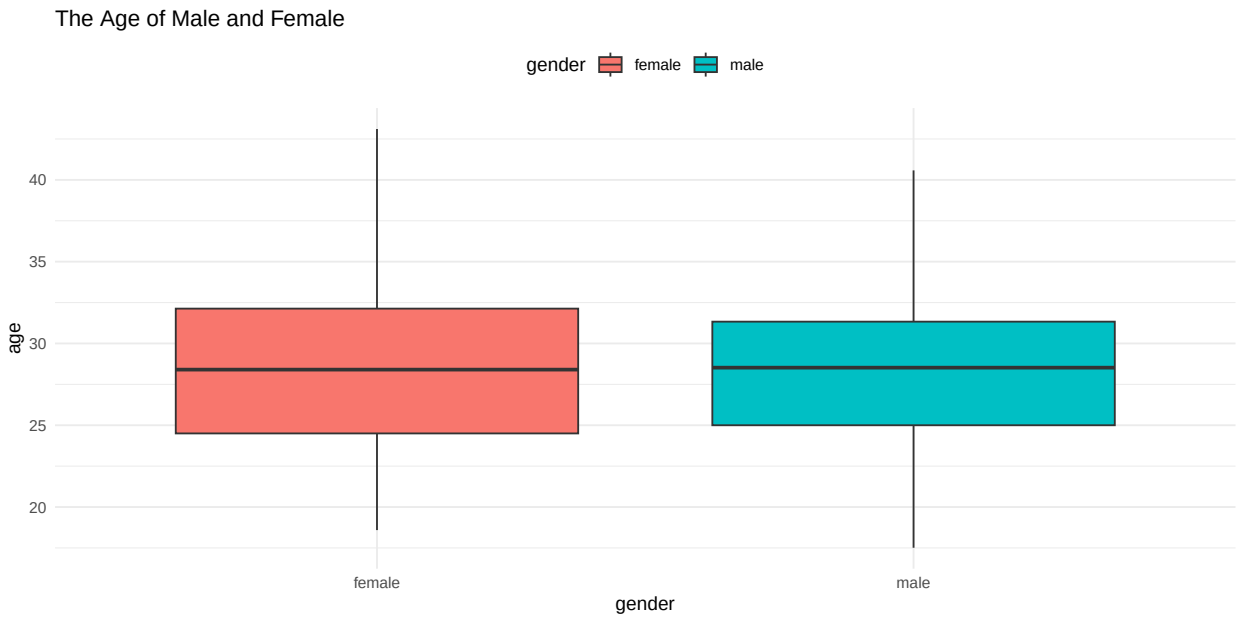
ID	gender	bodyweight	age	weightlifted
subj1	male	104.4	24.50	171.0
subj2	female	85.8	27.70	120.1
subj3	female	70.7	34.67	105.7
subj4	female	93.0	19.16	122.3
subj5	male	107.8	24.56	178.3

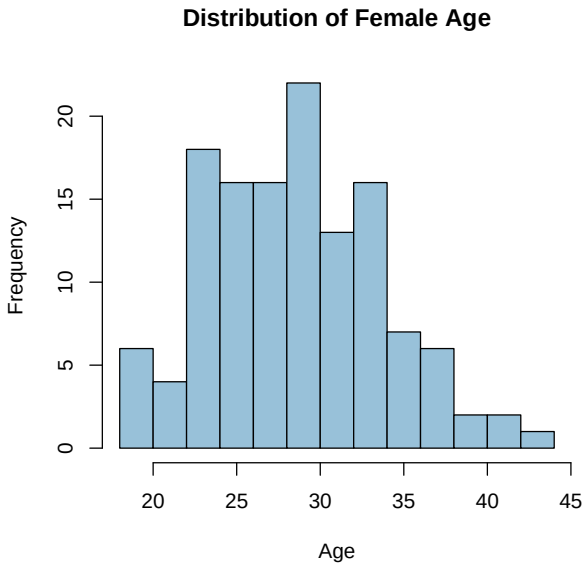
Table 3: Descriptive Statistics

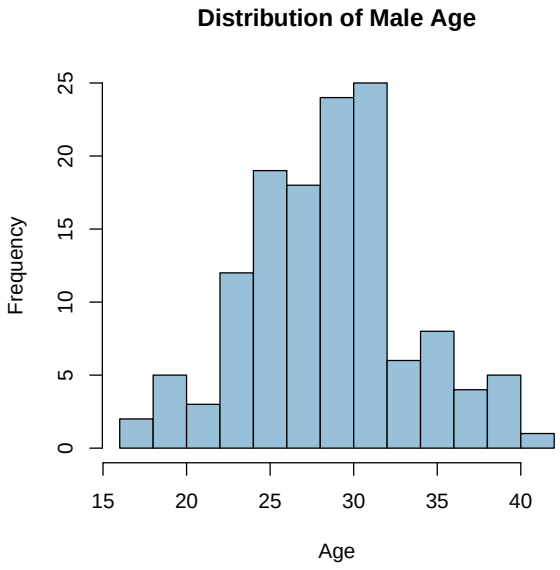
ID	gender	bodyweight	age	weightlifted
Length:261	Length:261	Min. : 63.30	Min. :17.51	Min. :105.7
Class :character	Class :character	1st Qu.: 85.80	1st Qu.:24.95	1st Qu.:115.6
Mode :character	Mode :character	Median :100.70	Median :28.47	Median :166.2
NA	NA	Mean : 97.54	Mean :28.49	Mean :146.7
NA	NA	3rd Qu.:109.80	3rd Qu.:31.48	3rd Qu.:176.9
NA	NA	Max. :124.00	Max. :43.11	Max. :187.3

Table 4: Descriptive Statistics of Two Samples Split By Gender

Variable	Count	Mean	StDev	StError
Female	129	28.5416	5.1224	0.4510
Male	132	28.4418	4.8615	0.4231







The comparative box plot illustrates that the means and ranges of the age of the two groups are mostly the same. In addition, the histograms show that the ages of female and male weightlifters were likely normally distributed. This allows us to establish the two independent sample t-tests. In this method, assumptions are stated about the average age of the two groups. The first assumption is that there is no difference between the age of female and male weight lifters. The alpha is obtained for the test to check the assumptions and achieve the result which is The first assumption will be rejected at an alpha level of $p < 0.05$. . On the contrary, there is not enough evidence to reject the first assumption when alpha $p > 0.05$. All analyses were performed using R, version 4.2.2.

Result 2

Research Question 1 : Is there any difference in the average age of female and male weightlifters?

`The test statistic is t = 0.1613668 with df = 259 .The p-value is p = 0.8719303 .`

Since the p-value is greater than 0.05, there is insufficient evidence to suggest that the average age of female weightlifters and the age of male weightlifters are different. In other words, their average age could be the same. To strengthen this claim, we can compute the 95% confidence interval of the two groups for comparison.

`The confidence interval of female weightlifters' age is (27.64917, 29.43393).`

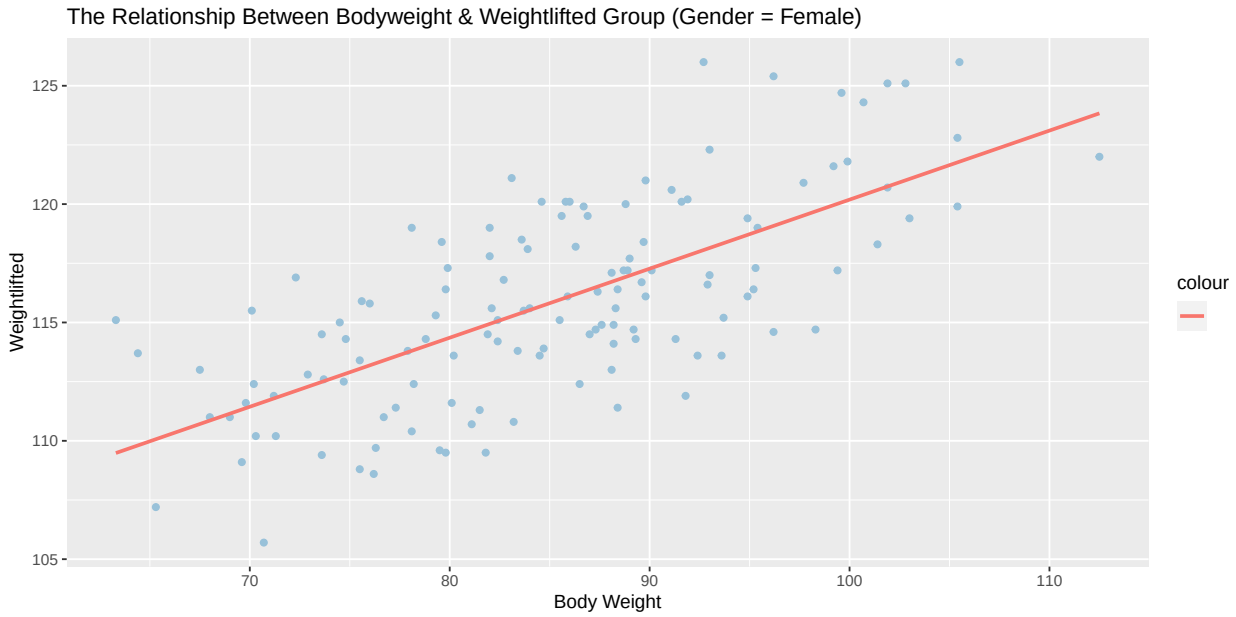
`The confidence interval of male weightlifters' age is (27.60476, 29.27888).`

We can conclude, on the basis of both the t-test and the Confidence Interval, that there is no difference between age of female and male weightlifters.

Research Question 2 : What is the relation between the body weight of weightlifters and the weight lifted?

1.Gender = Female: In fitting a linear model (i.e., the least-squares regression line) we must check several assumptions (occasionally called "conditions") to determine if a linear model is reasonable:

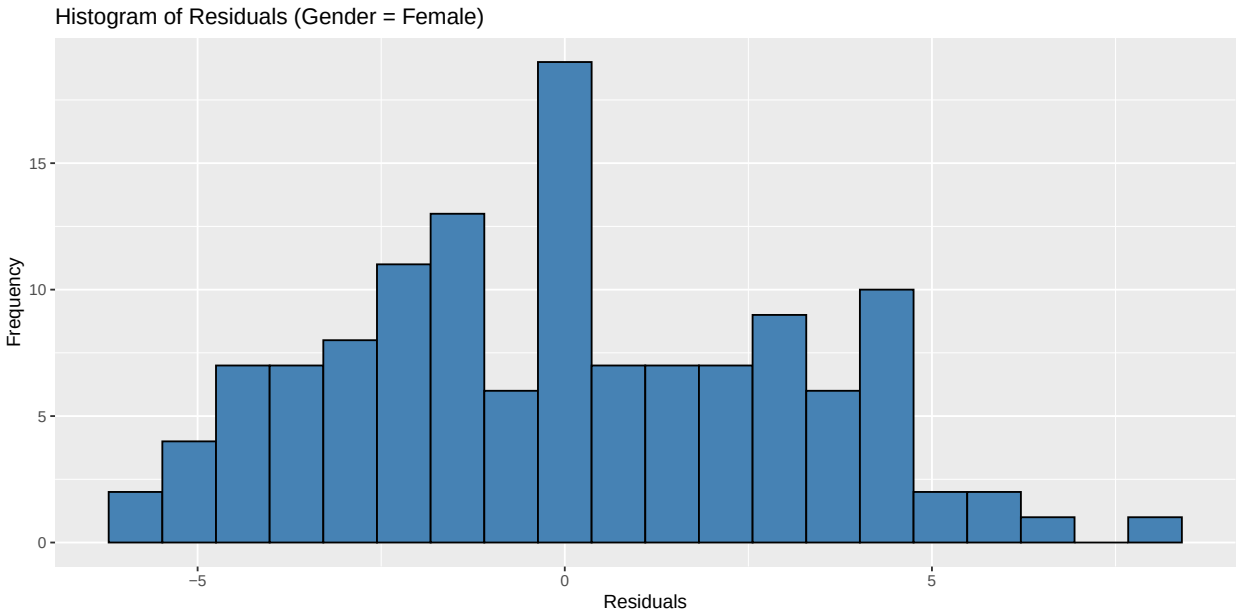
Check Assumption 1 (Scatter Plot) :



The graph indicates that there is a linear relationship between bodyweight and weightlifted variables.

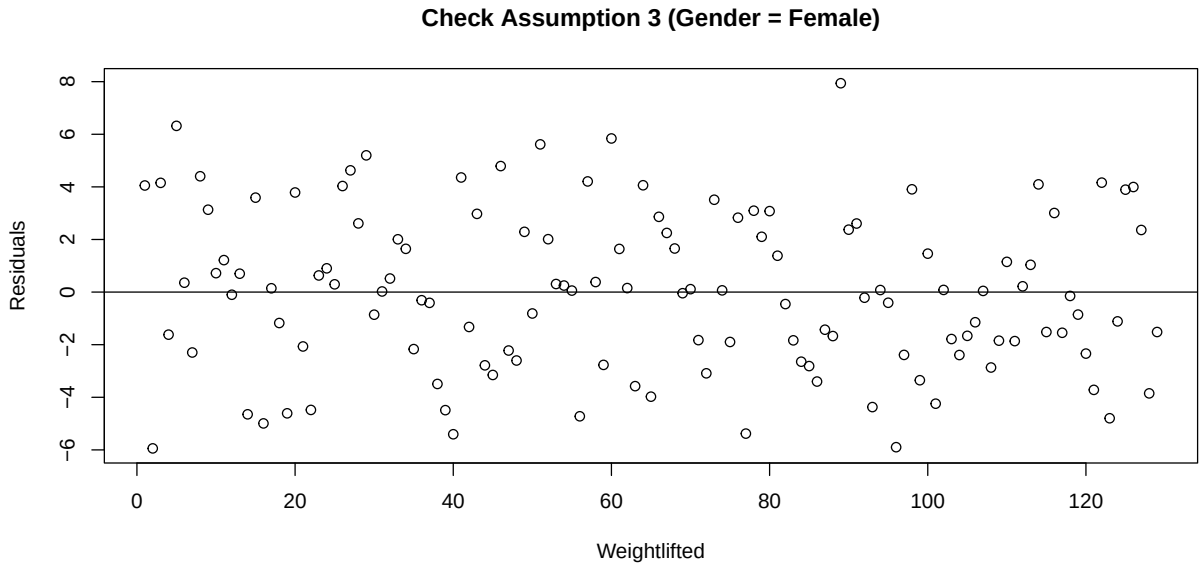
Check Assumption 2 (Residuals Are From A Normal Distribution) : Before checking the residuals' distribution, I have to fit the linear model of relationship between two variables.

```
Linear Model show relationship between bodyweight and weightlifted variables(Gender = Female) :
 weightlifted =  91.02183 *bodyweight +  0.2916705
```



The graphs show that residuals is normally distributed.

Check Assumption 3: Residuals Have A Constant Standard Deviation



This residuals plot give us no indication that the assumptions of model are false.

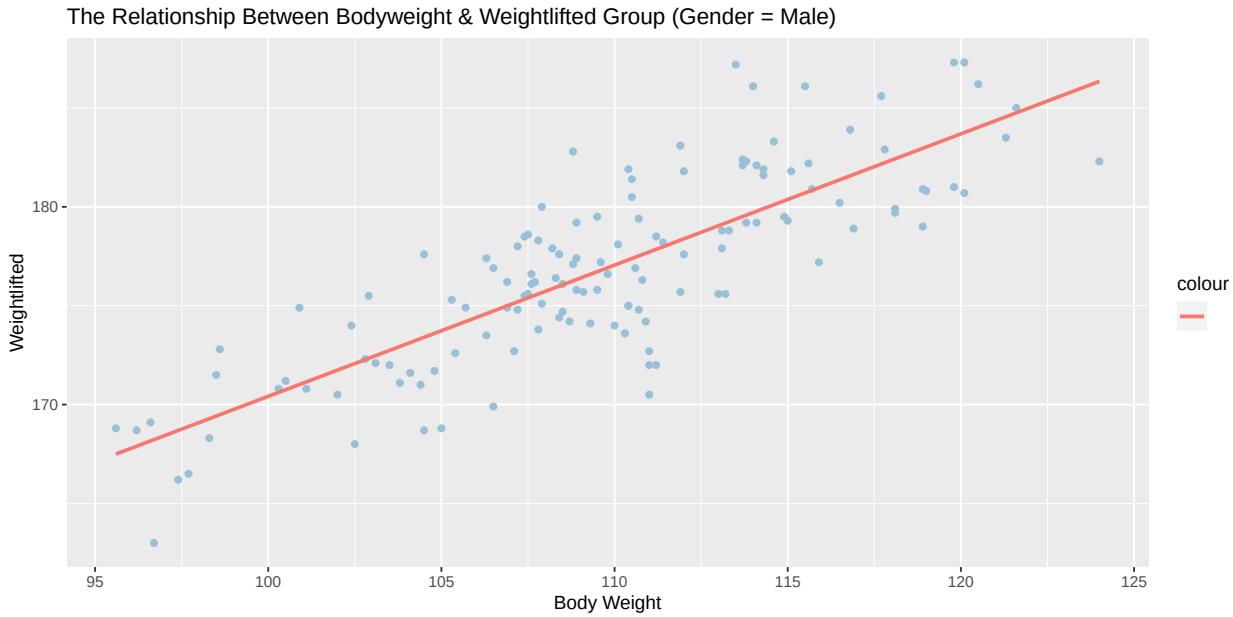
Hypothesis Test (Gender = Female) : In this step, the hypothesis about the relation between bodyweight and weightlifted variables will be tested. The first hypothesis states that there is no relationship between two above-mentioned variables. The t statistic and the alpha (p-value) are the tools to assess the hypothesis and draw the conclusion. This is the result of the analysis.

```
Linear Model (Gender = Female) : Weightlifted =  0.2916705 *bodyweight +  91.02183 .
 The t-stat =  0.02691917  and the p-value =  8.956537e-20  < 0.05.
```

Base on the p-value, we can reject the first hypothesis and conclude that there is a statistically significant positive linear relationship between the body weight and the maximum weight the athlete can lift. For each extra kilogram in weight, the maximum weight the athlete can lift increase by 0.2916 kilogram, on average.

2.Gender = Male

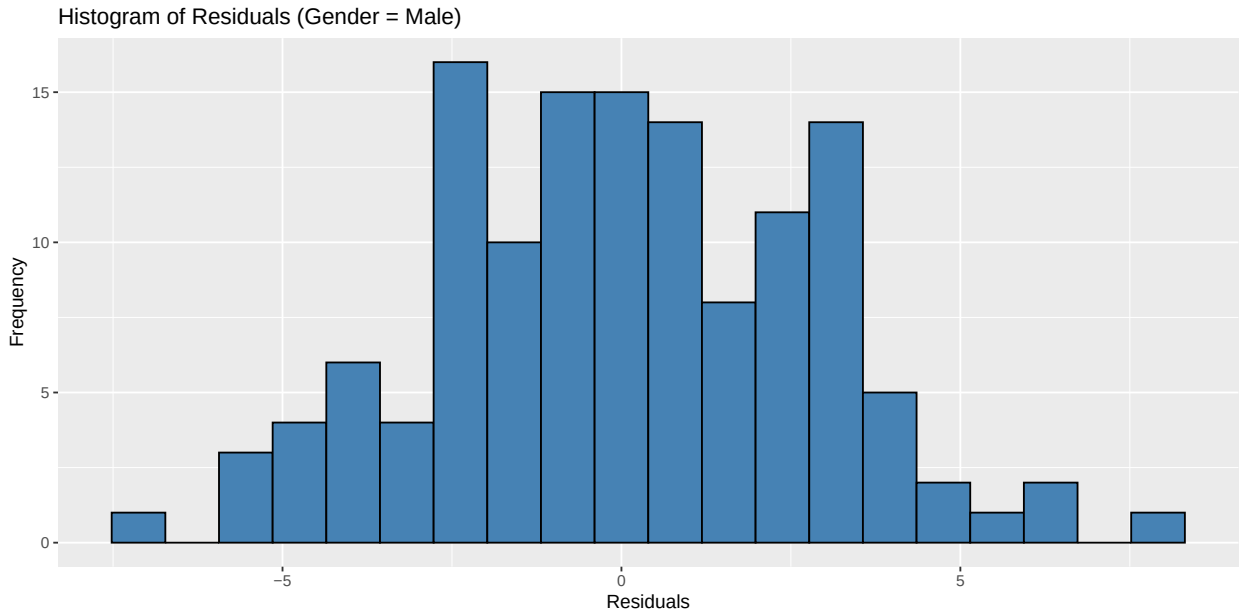
Check Assumption 1: Scatter Plot



Similarly, the graph for male gender show the linear relationship between two variables.

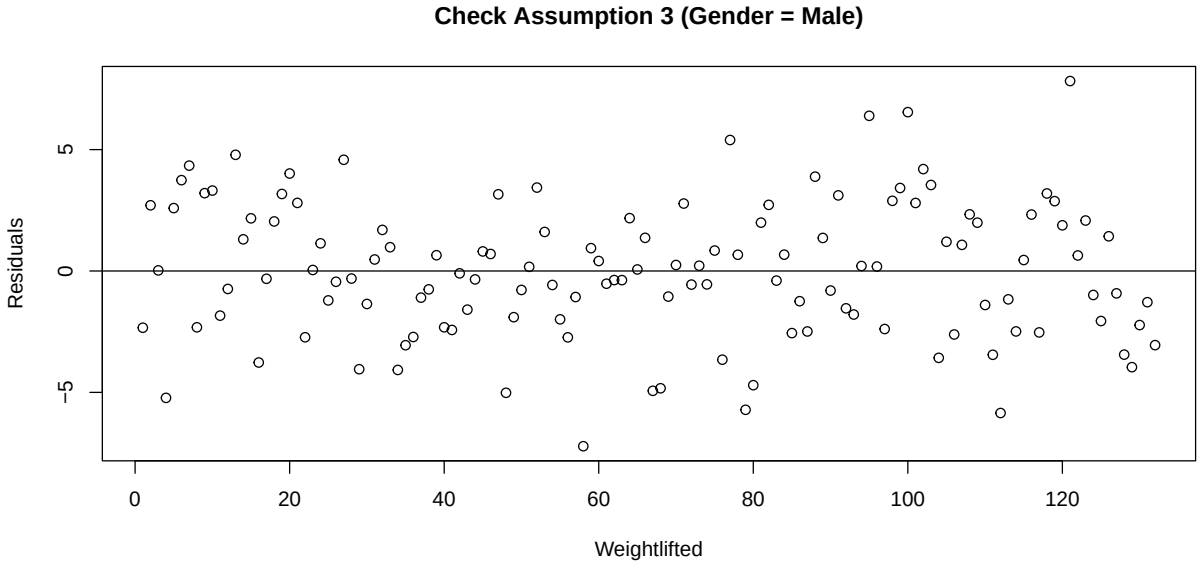
Check Assumption 2: Residuals Are From A Normal Distribution Before checking the residuals' distribution, I have to fit the linear model of relationship between two variables.

```
Linear Model show relationship between bodyweight and weightlifted variables(Gender = Male) :
 weightlifted =  104.0448 *bodyweight +  0.66372
```



The distribution of residuals are normal distribution.

Check Assumption 3: Residuals Have A Constant Standard Deviation



The residual plot indicates the residuals have a constant standard deviation. #### Hypothesis Test (Gender = Male):

Similarly, in this step, for male athletes, the hypothesis about the relation between bodyweight and weightlifted variables will be tested. The first hypothesis states that there is no relationship between two above-mentioned variables. The t statistic and the alpha (p-value) are the tools to assess the hypothesis and draw the conclusion. This is the result of the analysis.

```
Linear Model (Gender = Male) : Weightlifted =  0.66372 *bodyweight +  104.0448 .
 The t-stat =  0.04106818  and the p-value =  6.780118e-33  < 0.05.
```

Conclusions and Further Discussion

I quantified the impact of body weight and age and sex differences for weightlifters from the sample of 261 participants. The answer to the first question stated at the beginning of this study is that there is no difference between the age of female and male weightlifters. Besides that, there is a linear relationship between the body weight of the athletes and the maximum weight they can lift regardless the gender. Such results may help to establish progression trajectories for talented athletes and inform coaches, athletes, and national governing bodies.