Course Project

Topic 1: 3-Way Diet Comparison using One-Way ANOVA

IE7280:Statistical Methods of Engineering
Professor Nasser Fard

Aditya Trivedi
Ravi Patel

trivedi.adi@northeastern.edu
patel.ravikh@northeastern.edu

Topic 1: 3-Way Diet Comparison using One-Way ANOVA

Overview:

The 3-Diet Comparison Dataset is an extensive set of data that shows how people lost weight both before and after using one of three different diets. The dataset records a number of variables that cover different facets of the subjects' physical state both prior to and following the dietary changes. This dataset is a useful tool for researching how well these three diets work to promote weight reduction.

Problem Statement:

This study's main goal is to determine and contrast how well three different diets work to promote weight reduction. Through dataset analysis, our goal is to determine if there are any noteworthy variations in the results attained by those who followed each of these diets. The dataset provides a thorough assessment of the changes in participants' weight and associated factors by combining measurements made before and after the dietary treatments.

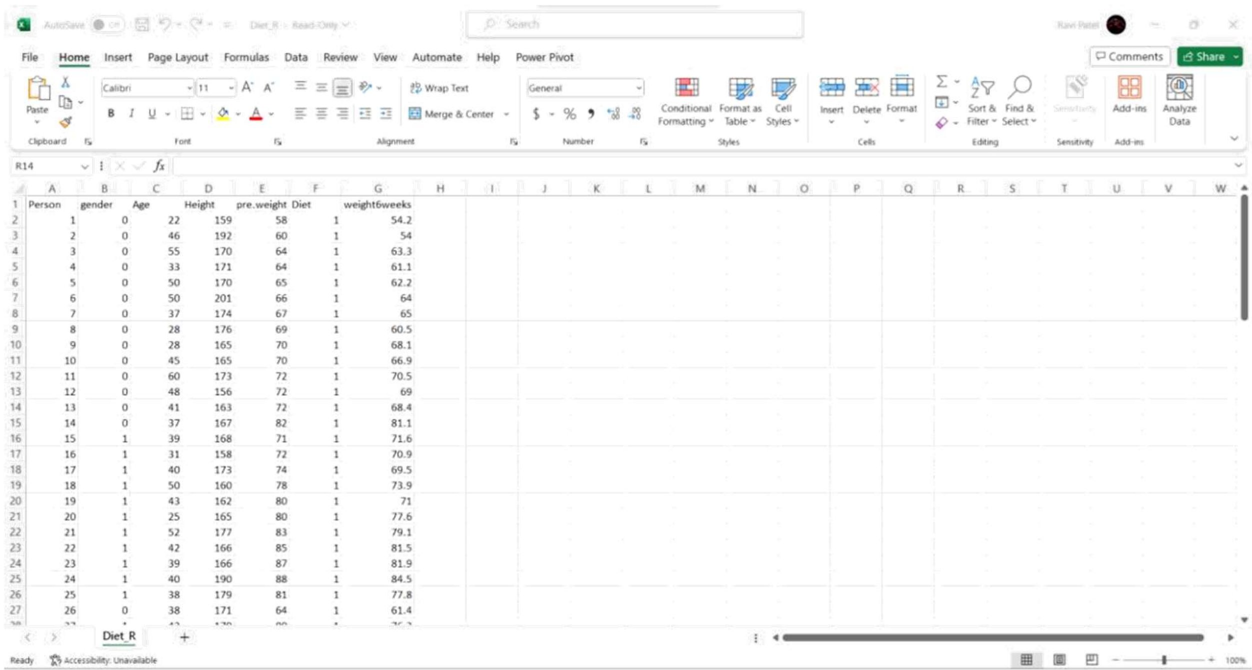
We want to use statistical techniques, such as ANOVA, to analyze the dataset methodically in order to ascertain whether there are significant differences in the three diets' results for weight reduction. The results of this research will provide insightful information on how effective each diet is in comparison, which may help those looking for the best weight loss plans.

The following particular research questions are directing this study:

- Are the three diets' respective results for weight loss statistically different?
- What is the average weight reduction and other key metrics of the three diets and how do they compare?
- Is it possible to pinpoint the diet or diets that work best for encouraging weight loss?

Dataset:

The Dataset snapshot is below, and it has 6 features.



Person	gender	Age	Height	pre.weight	Diet	weight6weeks
1	0	22	159	58	1	54.2
2	0	46	192	60	1	54
3	0	55	170	64	1	63.3
4	0	33	171	64	1	61.1
5	0	50	170	65	1	62.2
6	0	50	201	66	1	64
7	0	37	174	67	1	65
8	0	28	176	69	1	60.5
9	0	28	165	70	1	68.1
10	0	45	165	70	1	66.9
11	0	60	173	72	1	70.5
12	0	48	156	72	1	69
13	0	41	163	72	1	68.4
14	0	37	167	82	1	81.1
15	1	39	168	71	1	71.6
16	1	31	158	72	1	70.9
17	1	40	173	74	1	69.5
18	1	50	160	78	1	73.9
19	1	43	162	80	1	71
20	1	25	165	80	1	77.6
21	1	52	177	83	1	79.1
22	1	42	166	85	1	81.5
23	1	39	166	87	1	81.9
24	1	40	190	88	1	84.5
25	1	38	179	81	1	77.8
26	0	38	171	64	1	61.4

Exploratory Data Analysis:

The main 5 features namely Gender, Age, Height, Before Weight, and Weight after 6 weeks were drawn from the dataset which has 6 features and 90 records.

Below is how we read our data in R, displayed the first 10 records of the data set and a gist of string of the data.

```
> diet <- read.csv("Diet.csv")
> diet1<-read.csv("Diet.csv")
> head(diet, 10)
   Person_ID Gender Age Height Before_Weight Diet Weight_6Weeks
1          1      0  22    159            58    1          54.2
2          2      0  46    192            60    1          54.0
3          3      0  55    170            64    1          63.3
4          4      0  33    171            64    1          61.1
5          5      0  50    170            65    1          62.2
6          6      0  50    201            66    1          64.0
7          7      0  37    174            67    1          65.0
8          8      0  28    176            69    1          60.5
9          9      0  28    165            70    1          68.1
10        10      0  45    165            70    1          66.9
> str(diet)
'data.frame':   90 obs. of  7 variables:
 $ Person_ID    : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Gender       : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Age          : int  22 46 55 33 50 50 37 28 28 45 ...
 $ Height       : int  159 192 170 171 170 201 174 176 165 165 ...
 $ Before_Weight: int  58 60 64 64 65 66 67 69 70 70 ...
 $ Diet         : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Weight_6Weeks: num  54.2 54 63.3 61.1 62.2 64 65 60.5 68.1 66.9 ...
```

From the data, we can see that Gender, Age, Height, and type of diet matters for the Weight after 6 weeks.

Below is the summary of the data, showing each column's statistics. We tried to see the exact number of people who took the 3 diets.

```
> summary(diet)
   Person_ID         Gender            Age            Height       Before_Weight         Diet
 Min.   : 1.00   Min.   :0.0000   Min.   :16.00   Min.   :141.0   Min.   : 58.00   Min.   :1
 1st Qu.:23.25   1st Qu.:0.0000   1st Qu.:33.00   1st Qu.:164.2   1st Qu.: 66.00   1st Qu.:1
 Median :45.50   Median :0.0000   Median :39.00   Median :170.0   Median : 72.00   Median :2
 Mean   :45.50   Mean   :0.4444   Mean   :39.08   Mean   :170.9   Mean   : 72.53   Mean   :2
 3rd Qu.:67.75   3rd Qu.:1.0000   3rd Qu.:45.75   3rd Qu.:175.8   3rd Qu.: 79.00   3rd Qu.:3
 Max.   :90.00   Max.   :1.0000   Max.   :60.00   Max.   :201.0   Max.   :103.00   Max.   :3
 Weight_6Weeks
 Min.   : 53.00
 1st Qu.: 61.65
 Median : 68.80
 Mean   : 68.72
 3rd Qu.: 75.03
 Max.   :103.00
> diet_counts <- table(diet$Diet)
> print(diet_counts)

 1  2  3
30 30 30
```

We calculated the Starting BMI, Ending BMI, and Change in BMI because we can get to know if the diet was helpful for the person or shall he/she should change the diet down the road to make it more effective to help them lose weight.

```
> diet1$BMI_Start <- diet1$Before_Weight / (diet1$Height / 100)^2
> diet1$BMI_End <- diet1$Weight_6weeks / (diet1$Height / 100)^2
> diet1$BMI_Change <- diet1$BMI_Start - diet1$BMI_End
> print(diet1)
   Person_ID Gender Age Height Before_Weight Diet Weight_6weeks BMI_Start  BMI_End BMI_Change
1          1      0  22    159            58    1          54.2  22.94213 21.43903  1.5031051
2          2      0  46    192            60    1          54.0  16.27604 14.64844  1.6276042
3          3      0  55    170            64    1          63.3  22.14533 21.90311  0.2422145
4          4      0  33    171            64    1          61.1  21.88708 20.89532  0.9917581
5          5      0  50    170            65    1          62.2  22.49135 21.52249  0.9688581
6          6      0  50    201            66    1          64.0  16.33623 15.84119  0.4950373
7          7      0  37    174            67    1          65.0  22.12974 21.46915  0.6605892
8          8      0  28    176            69    1          60.5  22.27531 19.53125  2.7440599
9          9      0  28    165            70    1          68.1  25.71166 25.01377  0.6978880
10        10      0  45    165            70    1          66.9  25.71166 24.57300  1.1386593
11        11      0  60    173            72    1          70.5  24.05603 23.55575  0.5011861
12        12      0  48    156            72    1          69.0  29.58580 28.35306  1.2327416
13        13      0  41    163            72    1          68.4  27.09925 25.74429  1.3549626
14        14      0  37    167            82    1          81.1  29.40227 29.07957  0.3227079
15        15      1  39    168            71    1          71.6  25.15590 25.36848 -0.2125850
16        16      1  31    158            72    1          70.9  28.84153 28.40090  0.4406345
17        17      1  40    173            74    1          69.5  24.72518 23.22162  1.5035584
18        18      1  50    160            78    1          73.9  30.46875 28.86719  1.6015625
19        19      1  43    162            80    1          71.0  30.48316 27.05380  3.4293553
20        20      1  25    165            80    1          77.6  29.38476 28.50321  0.8815427
21        21      1  52    177            83    1          79.1  26.49303 25.24817  1.2448530
22        22      1  42    166            85    1          81.5  30.84628 29.57614  1.2701408
23        23      1  39    166            87    1          81.9  31.57207 29.72129  1.8507766
24        24      1  40    190            88    1          84.5  24.37673 23.40720  0.9695291
25        25      1  38    179            81    1          77.8  25.28011 24.28139  0.9987204
```

We made Change BMI for each diet.

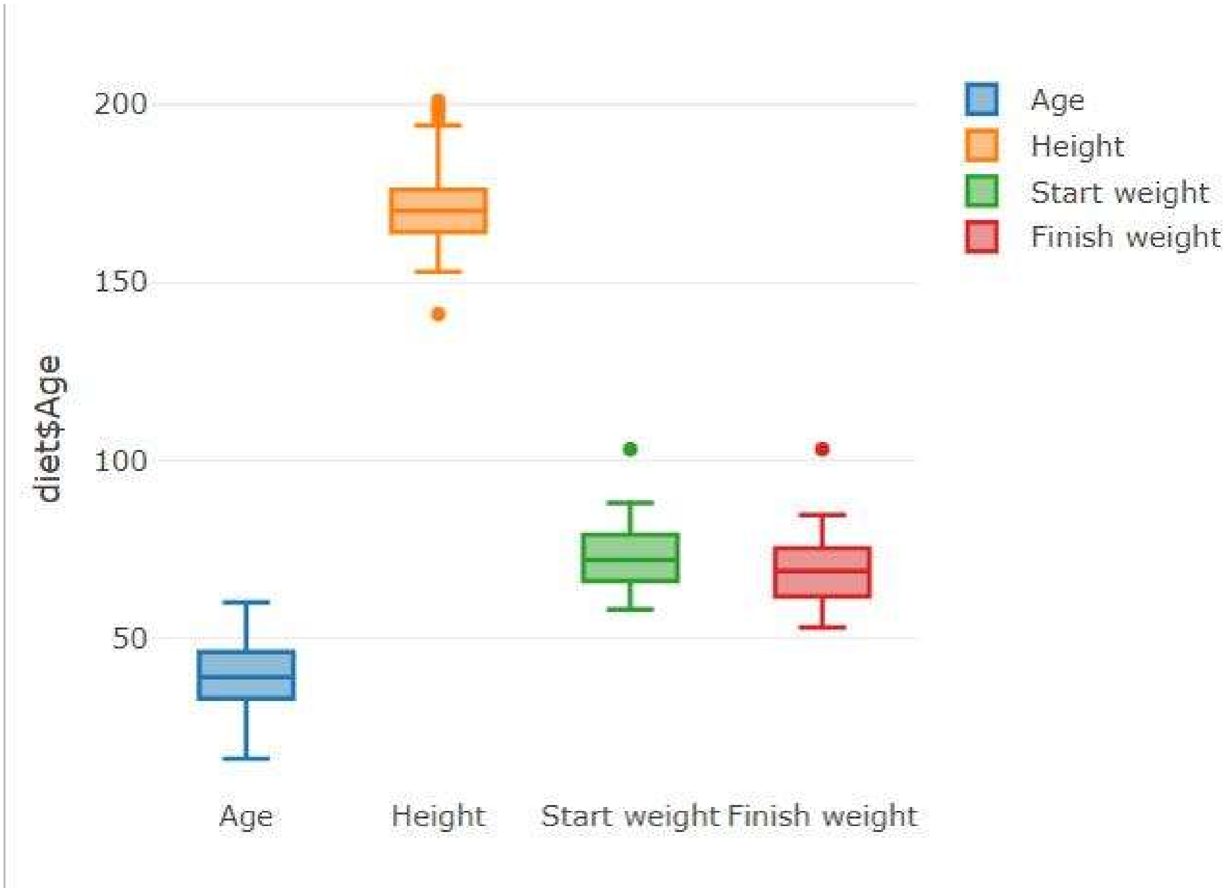
```
> diet_1 <- diet1[diet1$Diet == 1, "BMI_Change"]
> diet_2 <- diet1[diet1$Diet == 2, "BMI_Change"]
> diet_3 <- diet1[diet1$Diet == 3, "BMI_Change"]
> print(diet_1)
 [1]  1.5031051  1.6276042  0.2422145  0.9917581  0.9688581  0.4950373  0.6605892  2.7440599
 [9]  0.6978880  1.1386593  0.5011861  1.2327416  1.3549626  0.3227079 -0.2125850  0.4406345
[17]  1.5035584  1.6015625  3.4293553  0.8815427  1.2448530  1.2701408  1.8507766  0.9695291
[25]  0.9987204  0.8891625  1.1859805  1.1755102  1.0380623  1.2420490
> print(diet_2)
 [1]  0.0000000  0.0000000 -0.6936187  0.6760411  0.6244261  1.4705379  2.4508946  0.1981768
 [9]  1.0162219  1.2028467  1.0886921  0.7527570  1.5426997  1.6852522  1.2730990 -0.1750639
[17]  1.1512842  0.6060453  1.5097876  0.9695291  2.0703125  0.4516952  2.0324438  2.0861120
[25]  2.0151005 -0.4320988  1.2981524  0.7949343  0.8851989  0.9917581
> print(diet_3)
 [1] 2.5711662 1.9607157 1.3448835 2.3808690 3.0468750 1.8906901 2.5593737 2.9968783 3.5209497
[10] 2.5259516 0.3114187 2.5990903 1.4021408 2.6912726 2.0284799 1.0405827 0.2808901 1.2701408
[19] 0.1670620 0.8937406 2.6840610 1.4044505 0.8360954 1.3086916 1.9705532 3.2987916 1.9918367
[28] 2.2892820 1.5589569 2.3612751
>
```

Calculating mean for BMI Change of each diet.

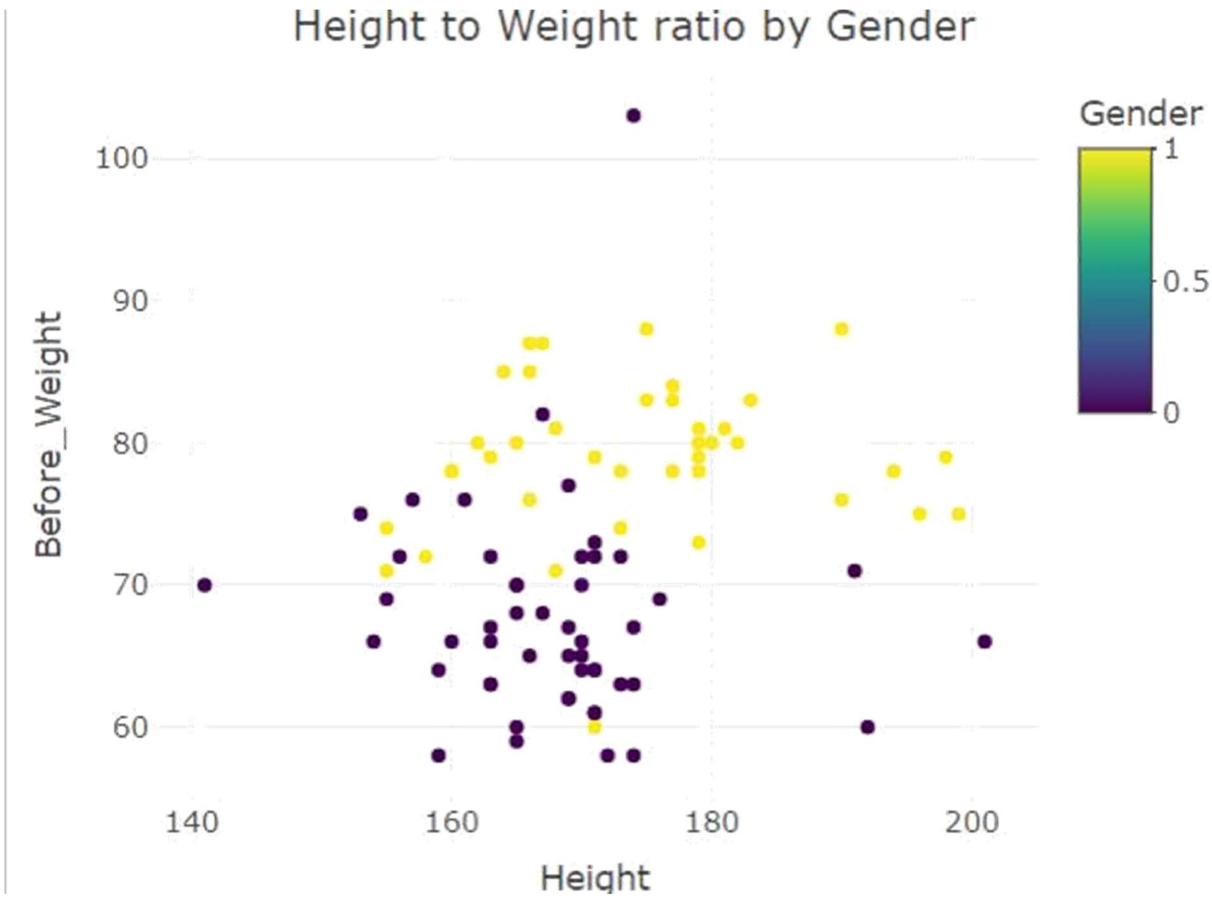
```
> Mean <- mean(diet1$BMI_Change)
> Mean
[1] 1.34134
```

<u>Data Visualization:</u>

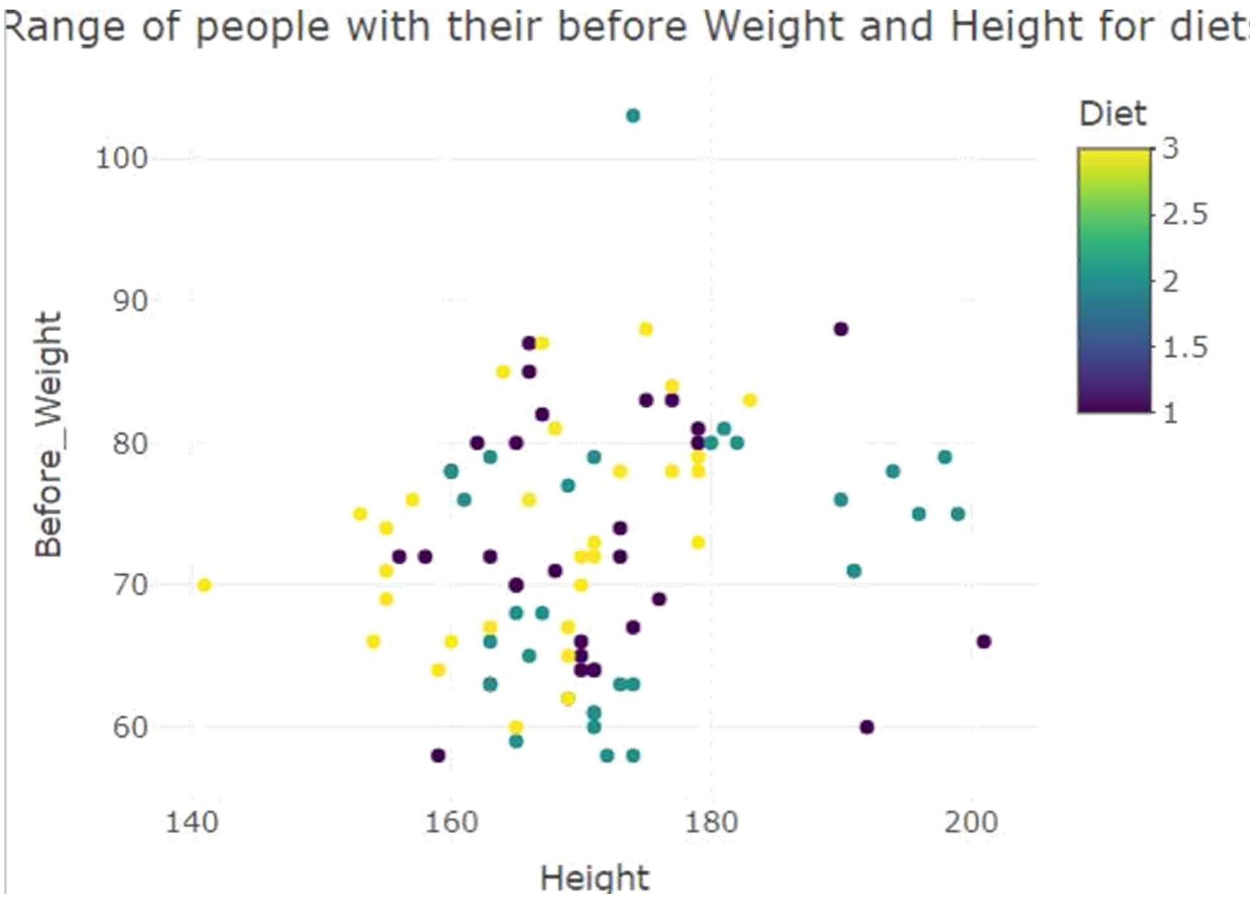
The boxplot below shows the comparison between the factor for diet i.e. Age, Height, Starting Weight, and Finish Weight.



The Scatter plot below shows the correlation between Height and Weight with respect to Gender



The Scatter plot below shows the correlation for the range of people with their Weight and Height for different diets.



The boxplot below shows comparison between 3 different diets.

Performing One Way ANOVA:

Calculating Sum of Squares Total for all the diets.

```
> diets <- list(diet_1, diet_2, diet_3)
> SST <- 0
> for (diet in diets) {
+   for (change in diet) {
+     SST <- SST + ((change - mean(unlist(diets)))^2)
+   }
+ }
> print(SST)
[1] 70.29682
```

Calculating Sum of Squares Within for all the diets.

```
> SSW <- function(diet1) {
+   Mean1 <- mean(diet1)
+   SSW <- sum((diet1-Mean1)^2)
+   return(c(Mean1, SSW))
+ }
>
> SSW_1 = SSW(diet_1)
> SSW_2 = SSW(diet_2)
> SSW_3 = SSW(diet_3)
> SSW_full <- SSW_1[2] + SSW_2[2] + SSW_3[2]
> print(SSW_full)
[1] 55.60724
```

Calculating Sum of Squares Between the difference between Sum of Squares Within and mean of BMI change of all diets.

```
> SSB <- 30*(abs(SSW_1[1] - Mean)^2 + abs(SSW_2[1] - Mean)^2 + abs(SSW_3[1] - Mean)^2)
> print(SSB)
[1] 14.68958
```

Calculating the F-Score of the data.

```
> fscore<-SSB/diet1_df/(SSW_full/diet1_df2)
> print(fscore)
[1] 11.75541
```

Finally, performing One Way ANOVA on the data.

```
> alldiet <- c(diet_1, diet_2, diet_3)
> group <- rep(1:3, each = length(diet_1))
> model <- aov(alldiet ~ group, data = diet1)
> summary(model)
            Df Sum Sq Mean Sq F value   Pr(>F)    
group        1   8.97   8.968   12.87 0.000548 ***
Residuals   88  61.33   0.697                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

As we can see in the One Way ANOVA Table, the p-value we received is 0.000548. As the p-value is smaller than the level of significance,0.05, thus we will be rejecting the H0, null hypothesis.

Summary:

We find that we need more people to experience weight loss with these 3 diets to make sure and get a piece of evidence that using these diets, people have a significant chance of losing weight than the people who won’t use these diets for their weight loss. Similarly, there is a moderation correlation between Gender and Diet the person takes. If a male takes the same diet, it seem that the diet give them better result as compared to female