Exam 1

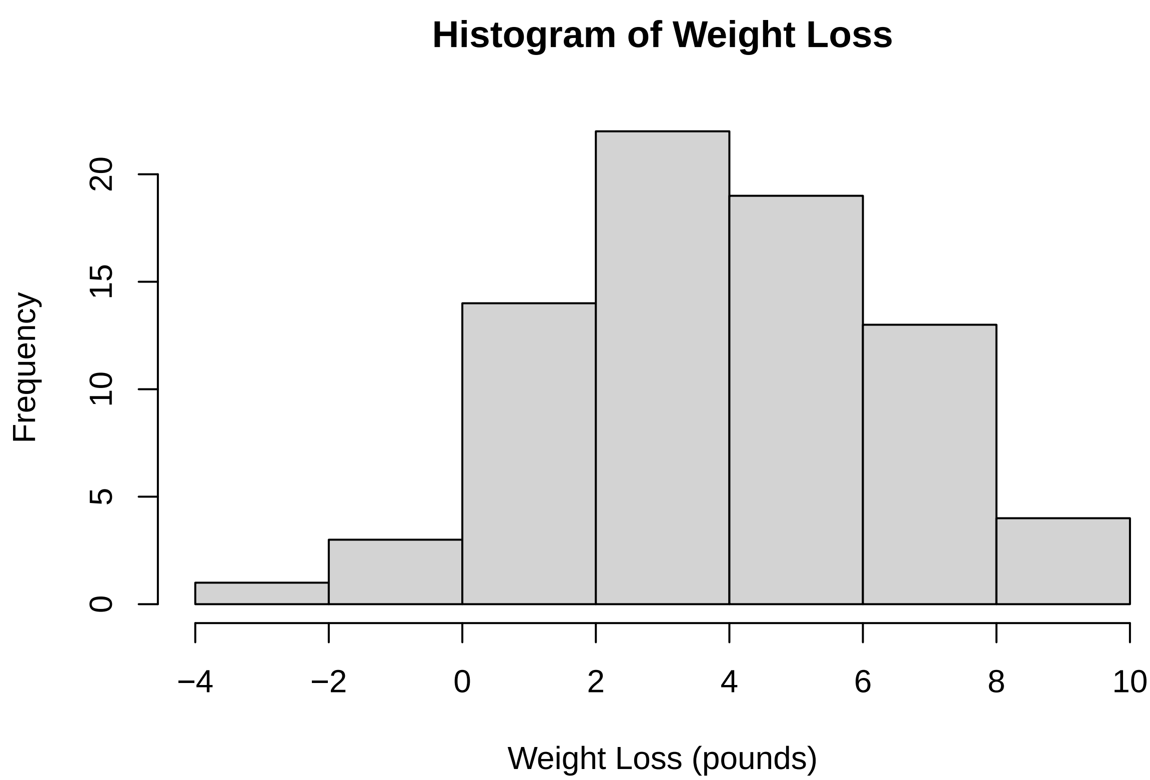
Aurora Travers

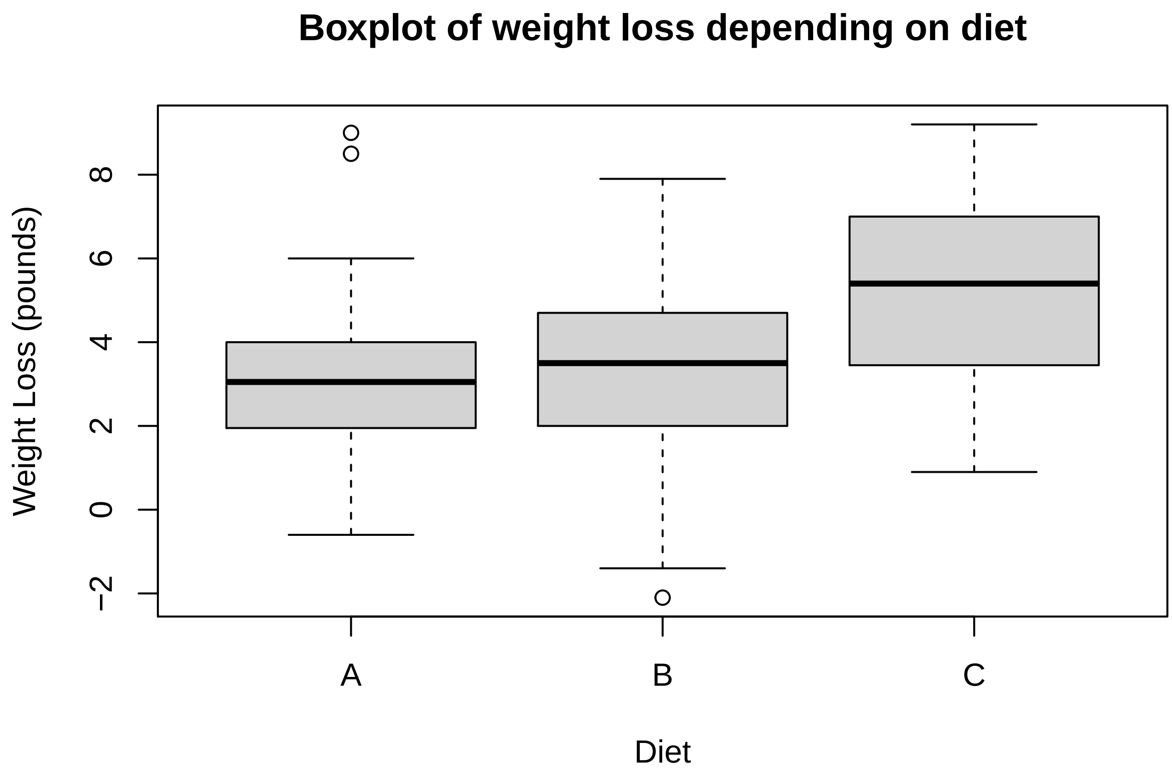
2022-08-18

I. Introduction

Obesity is a big problem in the world, and it is even bigger problem in the US. As of 2022, 42.4% of adults in the US is suffering because of obesity. Since obesity is closely linked with severe health issues such as heart diseases, it is crucial to solve this problem. To do so, we want to see which diet, out of A, B, and C, would be the most effective for losing weight. To approach this question, I would use single factor ANOVA.

II. Data Summary





The histogram suggests the data is approximatly normal since there is no skewness. Also, the most common weight loss was between 2 and 4 pounds.

The boxplot suggests the mean of diet A is the lowest and mean of diet C is the highest. Diet B has the largest interquartile range. There are two outliers in diet A and one outlier in diet B.

```
##   Diet Loss.mean  Loss.std Loss.samp_size
## 1    A  3.300000  2.240148      24.000000
## 2    B  3.268000  2.464535      25.000000
## 3    C  5.233333  2.247734      27.000000
```

The summary values are suggested above. The standard deviation of all three diets are similar, and the mean of diet A and B are similar. Diet C has the biggest mean, suggesting it could be the most effective diet.

III. Diagnostics

Since we are going to perform a single factor ANOVA, the assumptions are as follows:

- $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$
- Y_{ij} random and independent
- minimized under constraint $\sum_i \gamma_i = 0$

I believe the assumptions are met because the data is normal, random, and independent.

To view the data as a whole, there isn't any extreme values that we need to exclude.

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## Diet         2   66.2   33.09   6.154 0.00339 **
## Residuals   73  392.6    5.38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Analysis of Variance Table
##
## Response: Loss
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Diet       2  66.18  33.091  6.1537 0.00339 **
## Residuals 73 392.55   5.377
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Call:
## lm(formula = Loss ~ Diet, data = diet)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.3680 -1.4420  0.1167  1.5667  5.7000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.3000     0.4733   6.972 1.18e-09 ***
## DietB        -0.0320     0.6627  -0.048  0.96162
## DietC         1.9333     0.6506   2.972  0.00401 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.319 on 73 degrees of freedom
## Multiple R-squared:  0.1443, Adjusted R-squared:  0.1208
## F-statistic: 6.154 on 2 and 73 DF,  p-value: 0.00339
```

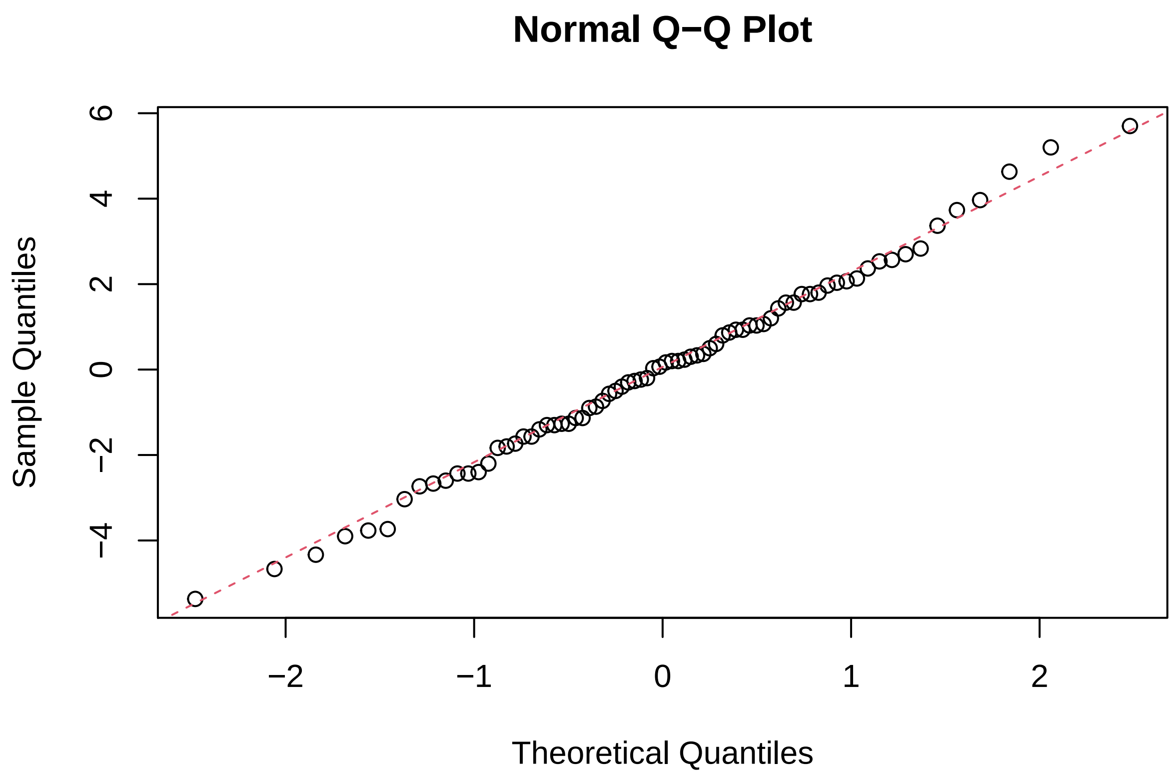
IV. Analysis

Our null hypothesis is $\mu_A = \mu_B = \mu_C$ when our alternative hypothesis is at least one μ_i is not the same.

Our F-value is 6.1537 and our p-value is 0.00339.

```
##           1           2           3           4           5           6
##  0.50000000  2.70000000 -2.60000000 -0.40000000 -0.50000000 -1.30000000
##           7           8           9          10          11          12
## -1.30000000  5.20000000 -1.40000000 -0.20000000 -1.80000000 -0.30000000
##          13          14          15          16          17          18
##  0.30000000 -2.40000000 -5.36800000 -1.26800000 -1.56800000  1.03200000
##          19          20          21          22          23          24
##  3.73200000 -2.66800000 -0.56800000  0.33200000 -0.26800000 -1.26800000
##          25          26          27          28          29          30
##  0.93200000  1.43200000  0.03200000 -3.76800000  1.76666667  0.36666667
##          31          32          33          34          35          36
## -1.83333333  1.56666667  2.56666667  0.16666667  1.56666667  1.96666667
```

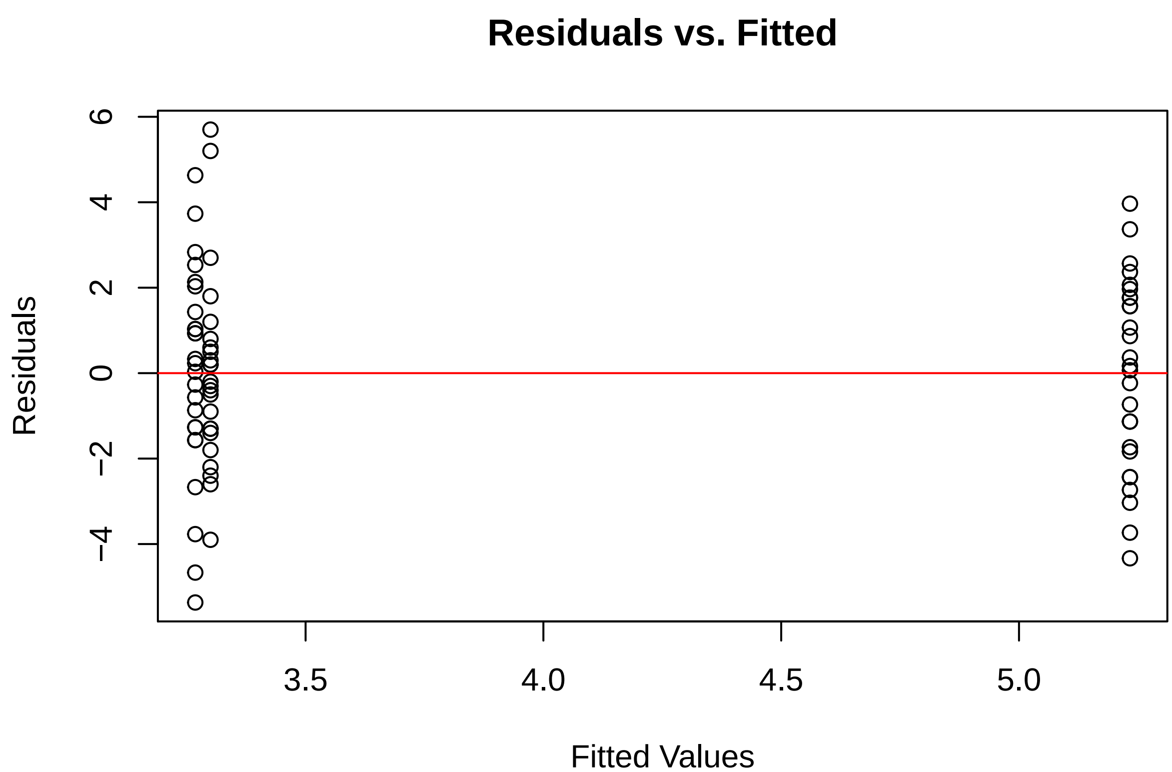
```
##           37          38          39          40          41          42
##   1.76666667  2.06666667 -4.33333333  2.36666667 -1.13333333  1.06666667
##           43          44          45          46          47          48
## -0.23333333 -3.90000000 -2.20000000  1.20000000  0.80000000  5.70000000
##           49          50          51          52          53          54
## -0.90000000  0.60000000  0.20000000  1.80000000  0.20000000  0.93200000
##           55          56          57          58          59          60
## -0.86800000  2.53200000  0.23200000  2.03200000 -1.56800000  2.13200000
##           61          62          63          64          65          66
##   2.83200000  4.63200000 -4.66800000  1.03200000 -2.73333333 -3.03333333
##           67          68          69          70          71          72
## -1.73333333 -3.73333333 -2.43333333  3.36666667 -0.73333333 -2.43333333
##           73          74          75          76
## -1.13333333  0.06666667  3.96666667  0.86666667
```




```
##
##  Shapiro-Wilk normality test
##
## data:  eij
## W = 0.99496, p-value = 0.9921
```

```
##        1        2        3        4        5        6        7        8
## 3.300000 3.300000 3.300000 3.300000 3.300000 3.300000 3.300000 3.300000
##        9       10       11       12       13       14       15       16
## 3.300000 3.300000 3.300000 3.300000 3.300000 3.300000 3.268000 3.268000
```

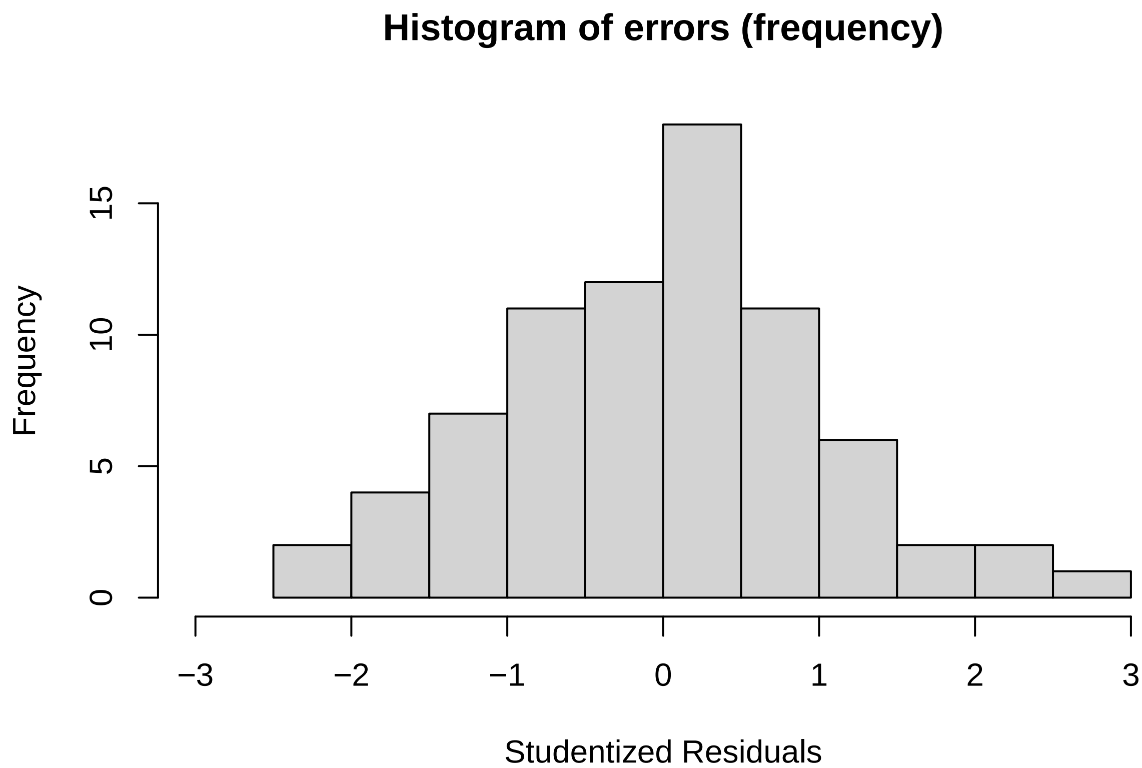
```
##       17       18       19       20       21       22       23       24
## 3.268000 3.268000 3.268000 3.268000 3.268000 3.268000 3.268000 3.268000
##       25       26       27       28       29       30       31       32
## 3.268000 3.268000 3.268000 3.268000 5.233333 5.233333 5.233333 5.233333
##       33       34       35       36       37       38       39       40
## 5.233333 5.233333 5.233333 5.233333 5.233333 5.233333 5.233333 5.233333
##       41       42       43       44       45       46       47       48
## 5.233333 5.233333 5.233333 3.300000 3.300000 3.300000 3.300000 3.300000
##       49       50       51       52       53       54       55       56
## 3.300000 3.300000 3.300000 3.300000 3.300000 3.268000 3.268000 3.268000
##       57       58       59       60       61       62       63       64
## 3.268000 3.268000 3.268000 3.268000 3.268000 3.268000 3.268000 3.268000
##       65       66       67       68       69       70       71       72
## 5.233333 5.233333 5.233333 5.233333 5.233333 5.233333 5.233333 5.233333
##       73       74       75       76
## 5.233333 5.233333 5.233333 5.233333
```




```
##
##   Brown-Forsythe Test (alpha = 0.05)
## -------------------------------------------------------------
##   data : Loss and Diet
##
##   statistic  : 6.1469
##   num df     : 2
##   denom df   : 71.87429
##   p.value    : 0.003432808
```

```
##
##   Result     : Difference is statistically significant.
## -------------------------------------------------------------

##           1           2           3           4           5           6
##  0.21881263  1.19280924 -1.14780691 -0.17502915 -0.21881263 -0.57000553
##           7           8           9          10          11          12
## -0.57000553  2.36134049 -0.61407351 -0.08750062 -0.79087675 -0.13125965
##          13          14          15          16          17          18
##  0.13125965 -1.05808302 -2.44155309 -0.55542880 -0.68761936  0.45172586
##          19          20          21          22          23          24
##  1.66226324 -1.17735485 -0.24837908  0.14513839 -0.11715394 -0.55542880
##          25          26          27          28          29          30
##  0.40784741  0.62763701  0.01398722 -1.67892257  0.77422373  0.16005167
##          31          32          33          34          35          36
## -0.80369702  0.68596612  1.13005258  0.07274049  0.68596612  0.86273127
##          37          38          39          40          41          42
##  0.77422373  0.90708802 -1.93998097  1.04061529 -0.49545964  0.46622421
##          43          44          45          46          47          48
## -0.10184028 -1.74174468 -0.96870726  0.52598342  0.35028194  2.60883315
##          49          50          51          52          53          54
## -0.39415640  0.26261358  0.08750062  0.79087675  0.08750062  0.40784741
##          55          56          57          58          59          60
## -0.37978254  1.11627271  0.10141442  0.89309311 -0.68761936  0.93756841
##          61          62          63          64          65          66
##  1.25125320  2.08486550 -2.10205947  0.45172586 -1.20486618 -1.34023991
##          67          68          69          70          71          72
## -0.75949757 -1.66022158 -1.07038837  1.49183991 -0.32027431 -1.07038837
##          73          74          75          76
## -0.49545964  0.02909530  1.76835410  0.37861300
```



Our Q-Q plot and Shapiro-Wilk normality test suggests that our data is normal. Our Residuals vs Fitted plot and Brown-Forsythe Test suggests that our variance is not constant.

```
## [1] 0.8962294
```

```
## [1] 2.403771
```

```
## [1] 4.196229
```

We are 95% confident that the true average of weight loss of diet A is between 2.403771 and 4.196229 pounds.

```
center <- 3.26800
stddev <- 2.464535
n <- 25
error <- qnorm(0.975)*stddev/sqrt(n)
error
```

```
## [1] 0.96608
```

```
lower_bound <- center - error
lower_bound
```

```
## [1] 2.30192
```

```
upper_bound <- center + error
upper_bound
```

```
## [1] 4.23408
```

We are 95% confident that the true average of weight loss of diet B is between 2.30192 pounds and 4.23408 pounds.

```
## [1] 0.8478346
```

```
## [1] 4.385498
```

```
## [1] 6.081168
```

We are 95% confident that the true average of weight loss of diet C is between 4.385498 and 6.081168 pounds.

```
## [1] 2.318935
```

According to https://homepage.univie.ac.at/robin.ristl/samplesize.php?test=anova, with a mean of 3.300000, 3.268000, 5.233333, standard deviation of 2.318935, alpha of 0.05, sample size of 24, 25, 27, we get a power of 0.8778.

V. Interpretation

Using $\alpha = 0.05$, we reject our null hypothesis. Thus, we conclude one of the diets differ in group mean. According to our 95% confidence interval and the group means, diet C is most effective in losing weight and diet A and B has a similar effect.

VI. Conclusion

Our group means are significantly different. Diet C is most effective in losing weight.

```
knitr::opts_chunk$set(echo = TRUE)
diet <- read.csv("loseit.csv")
hist(diet$Loss, xlab = "Weight Loss (pounds)", main = "Histogram of Weight Loss")
boxplot(Loss~Diet, data = diet, xlab = "Diet", ylab = "Weight Loss (pounds)", main = "Boxplot of weight
aggregate(Loss~Diet, data = diet, FUN = function(x) {c(mean = mean(x), std = sd(x), samp_size = length(
Q <- quantile(diet$Loss, probs=c(.25, .75), na.rm = FALSE)
iqr <- IQR(diet$Loss)
up <-  Q[2]+1.5*iqr
low<- Q[1]-1.5*iqr
eliminated<- subset(diet, diet$Loss > (Q[1] - 1.5*iqr) & diet$Loss < (Q[2]+1.5*iqr))
anova_diet <- aov(Loss~Diet, data = diet)
summary(anova_diet)
anova(anova_diet)
model <- lm(Loss~Diet, data = diet)
summary(model)
eij = residuals(model)
eij
qqnorm(eij, main = "Normal Q-Q Plot")
qqline(eij, col = 2, lty = 2)
```

```
shapiro.test(eij)
fitted_val = fitted(model)
fitted_val
plot(fitted_val, eij, main = "Residuals vs. Fitted", ylab = "Residuals", xlab = "Fitted Values")
abline(h=0, col = 'red')
library(onewaytests)
bf.test(Loss~Diet, data = diet)
library(MASS)
eij.stud = studres(model)
eij.stud
hist(eij.stud, main = "Histogram of errors (frequency)", xlab = "Studentized Residuals", xlim = c(-3, 3
center <- 3.300000
stddev <- 2.240148
n <- 24
error <- qnorm(0.975)*stddev/sqrt(n)
error
lower_bound <- center - error
lower_bound
upper_bound <- center + error
upper_bound
center <- 3.26800
stddev <- 2.464535
n <- 25
error <- qnorm(0.975)*stddev/sqrt(n)
error
lower_bound <- center - error
lower_bound
upper_bound <- center + error
upper_bound
center <- 5.233333
stddev <- 2.247734
n <- 27
error <- qnorm(0.975)*stddev/sqrt(n)
error
lower_bound <- center - error
lower_bound
upper_bound <- center + error
upper_bound
variance<-((2.240148)^2*(24-1)+(2.464535)^2*(25-1)+(2.247734)^2*(27-1))/(24+25+27-3)
sqrt(variance)
```