**PROJECT ON DIET AND WEIGHT ANALYSIS**

**TEAM NO**: 13

**TEAM MEMBERS:**

1. GENGA K(2033012)
2. MITHRAADHARSHINI M K(2033021)
3. POOJA M(2033024)

**LANGUAGE USED:** R

**DIET AND WEIGHT ANALYSIS**

**Introduction:**

Health plays role in humans to run their daily life. It determines the physical and mental state of a body. To maintain the health in a good condition various techniques are followed such as exercises ,yoga ,diet etc.In this project, Diet is taken for analysis and it is performed using various statistical tools in R.

**Problem statement**

* Diet is effective to loss weight by taking healthy food in regular intervals which in turn maintains the health of a body.
* There are different diet types followed.
* People are confused to choose which diet type is more effective

**Objective**

* Here 3 diet types are taken for analysis
* To analyse the weight loss in 3 diet types.
* To provide a solution for the effective diet type
* To analyse the results of diet types depending on gender to make the results more effective

**Dataset used**

* Data for three diet types are collected.
* Dataset is collected from Kaagle.
* It consists of 7 features and 76 records.
* Dataset is analysed with various tools in R

**Previously used techniques**

1.Descriptive Statistics

2.One way anova

3.Two way anova

4.Kruskal Wallis rank sum test

5.Non parametric Test

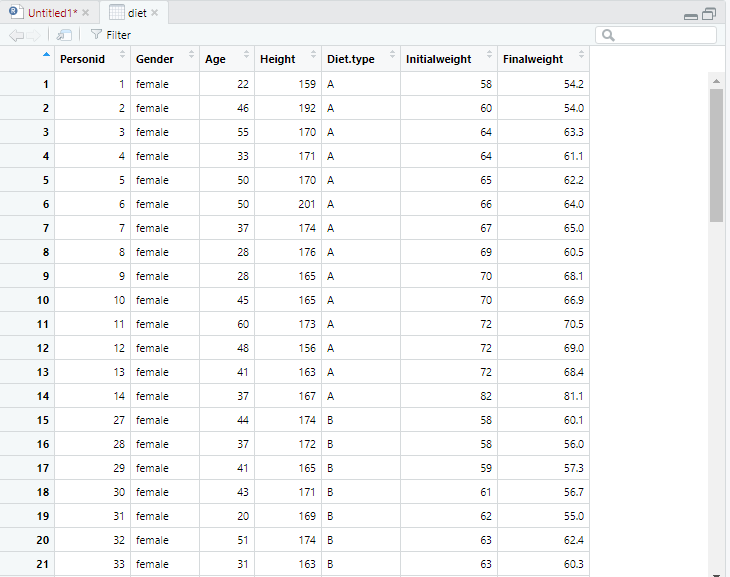
6.Data Visualisation(Box plot)

**Updated techniques**

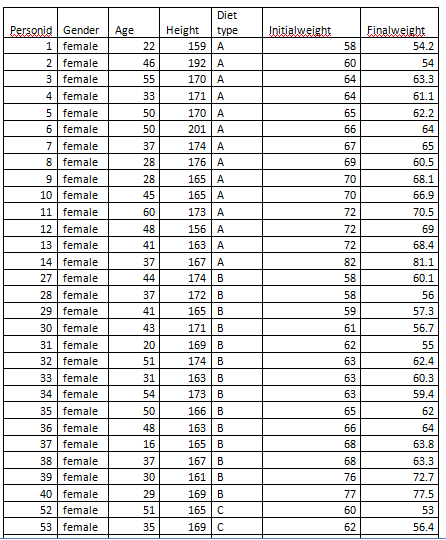
1. Linear Regression and Scatter plot
2. Time series Analysis
3. Cumsum
4. Markov Chain Monte Carlo
5. Auto Regression
6. Auto Correlation

**Importing dataset**



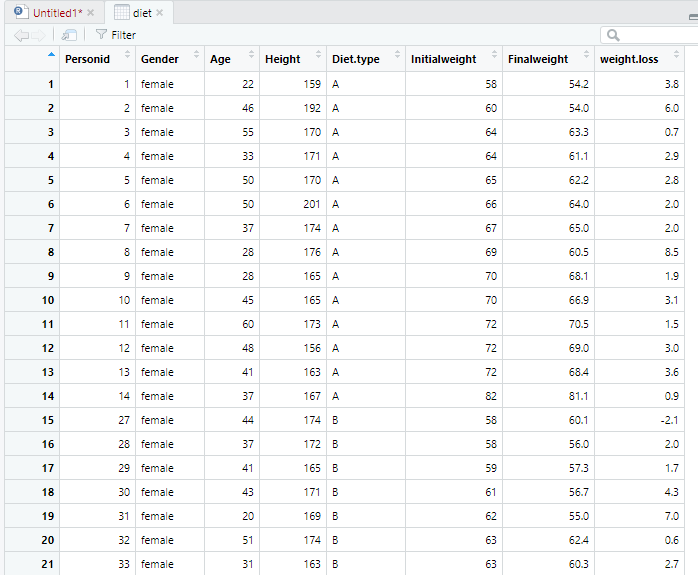


**Dataset**



**Calculating weight loss**





**Categorizing and store the data**

**factor()**

* Factors are used to represent the categorical data.
* Factors can be ordered and are an important for statistical analysis and for **plotting.**



**levels():**

* Factors can only contain pre-defined set values, known as **levels**
* R always sort levels in ascending order.

**Boxplot (weightloss vs diet type)**

* Boxplot are a measure of how well the data are distributed in a dataset.
* This graph represents the minimum, maximum, median, first quartile, second quartile ,outliers in the dataset.

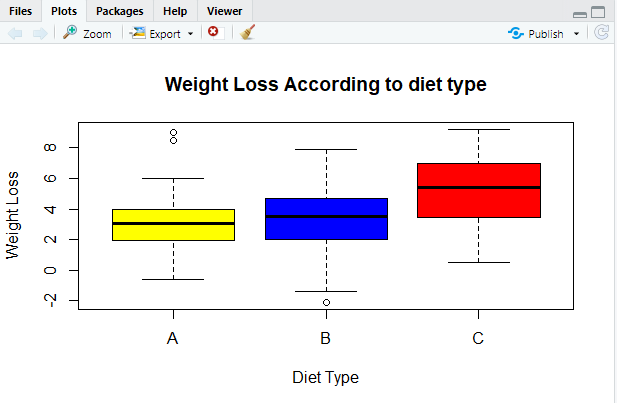
boxplot (diet$weight.loss~diet$Diet.type,

ylab ="Weight Loss",

xlab="Diet Type",

col=c("yellow","blue","red"),

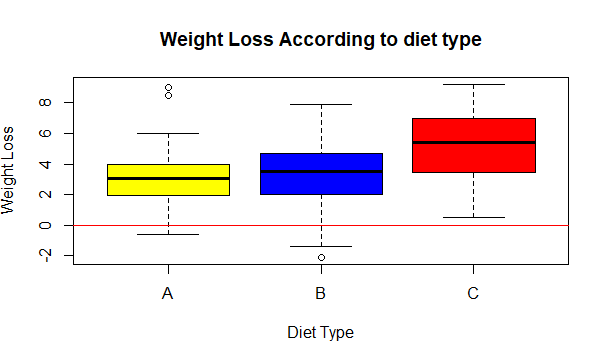
main="Weight Loss According to diet type")



**abline():**

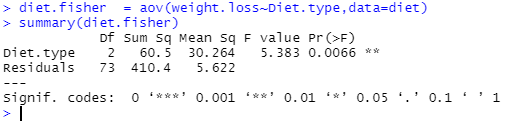
* It is used to add a one or more straight lines to a graph.
* It is used to add vertical ,horizontal or regression lines to plot.





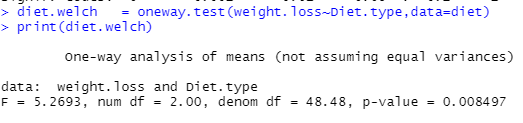
**One way anova**

* It is used to determine whether there is any significant difference between the average weight loss in 2 diet types.
* Here p<0.05 then there exists the significant differences between the groups.



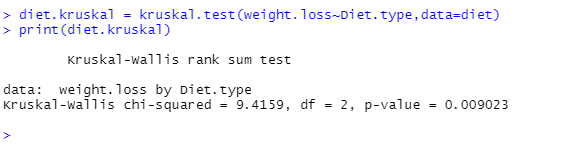
**One way analysis of means**

* Test whether two or more samples from normal distributions have the same means.
* The variances are not necessarily assumed to be equal.



**Kruskal-Wallis rank sum test**

* It is used to determine if there are statistically significant differences between 2 or more groups of independent variable on a continuous dependent variable.
* It is the nonparametric approach to one way anova**.**

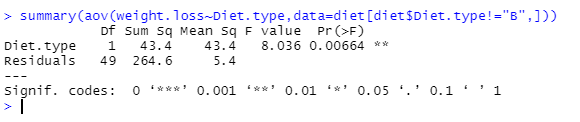


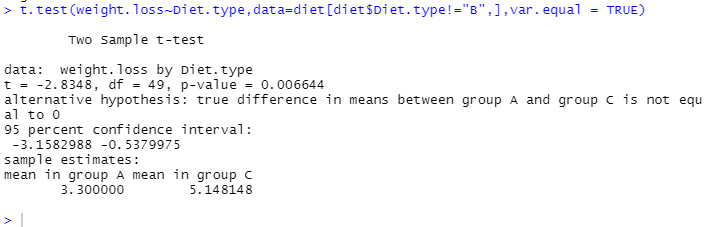
**INFERENCE**

* Here the p value is less than the significance level 0.05, we can conclude that there are significant differences between the diet types.

**Two sample t-test**

* The two sample T test also known as the independent sample T test is a method used to test whether the unknown population means of two groups are equal or not.





**Inference:**

* Here we use two sample T test for examining the difference in means for 2 population
* The mean values significantly differ from sample A to sample C
* Here sample A has a mean of 3.300000 and sample C has a mean of 5.148148

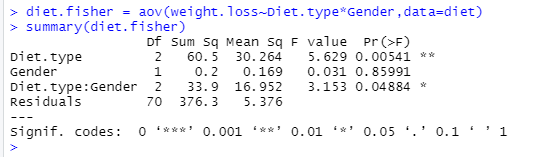
**Two way Anova**

The two way anova is the extension of one way

anova that examines the influence of two

different categorical independent variable on one

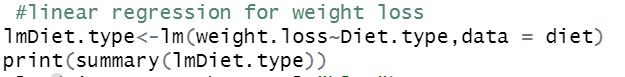
continuous dependent variable

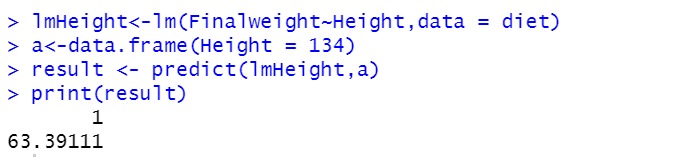


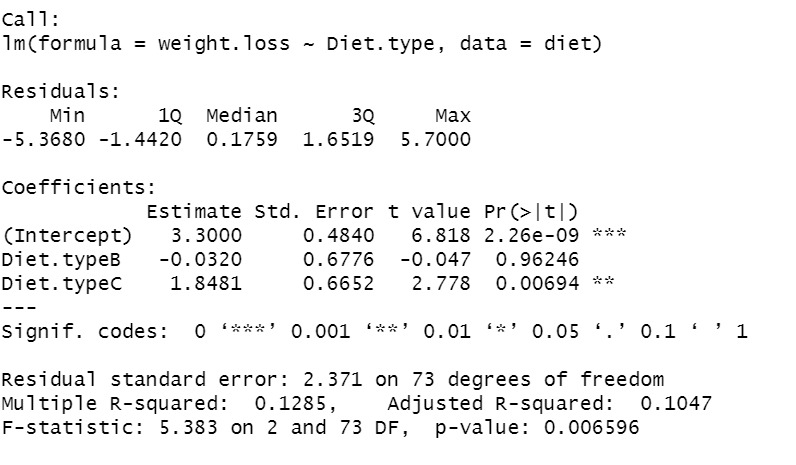
**Inference**

* Here in our project we use two way anova to reduce the occurrence of errors to certain extent
* When we add one more independent variable, it will reduce the residual values to a certain extent
* Hence by using two way we can reduce the variations that are caused by the independent variable

**Linear regression**





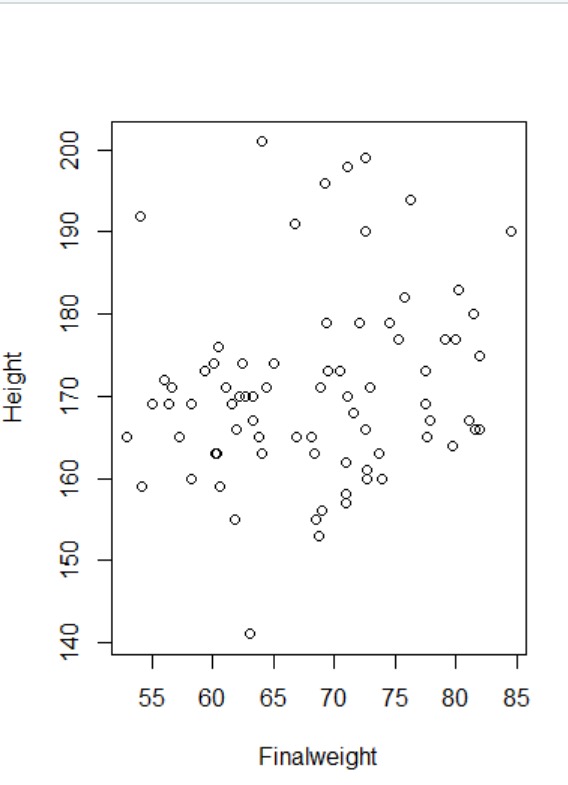


**Inference:**

* From this , the final weight of a person after the diet is mainly dependent on the height of a person.
* By using diet C one can able to reduce upto 3kg.
* By analysing the previous data, we can able to predict the weight by height.
* When the person height is 134cm,by using linear regression we predict the weight of the person is 63.39kg

**Scatter plot**

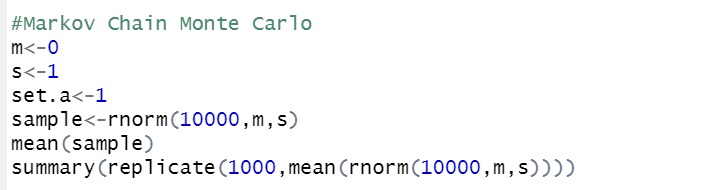
WhatsApp Image 2021-12-23 at 3.32.49 PM.jpeg

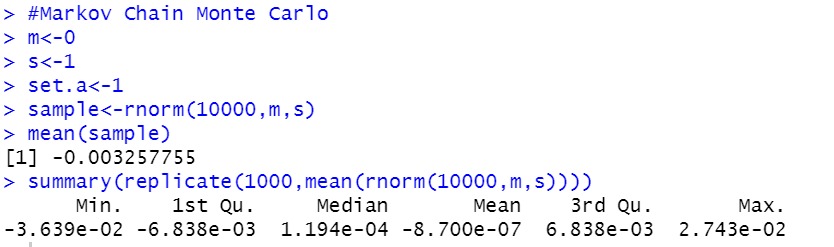


**Inference**

* Here the interpretation of the linear regression is visualized by the scatter plot .
* The People who are practicing diet having weight ranges between 55-80 mostly lies on the height range160 to 175.

**Markov Chain Monte Carlo**

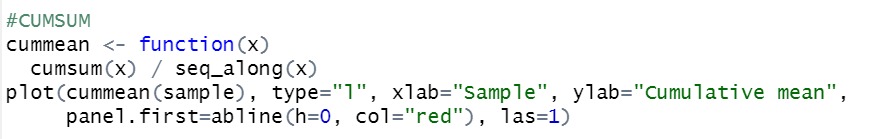


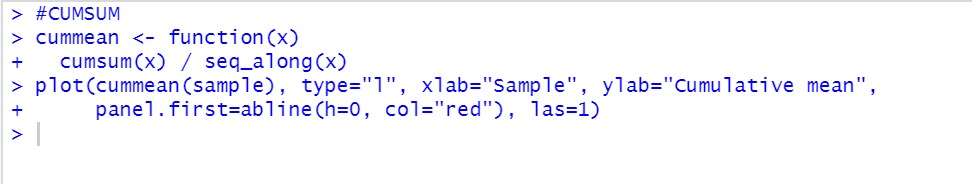


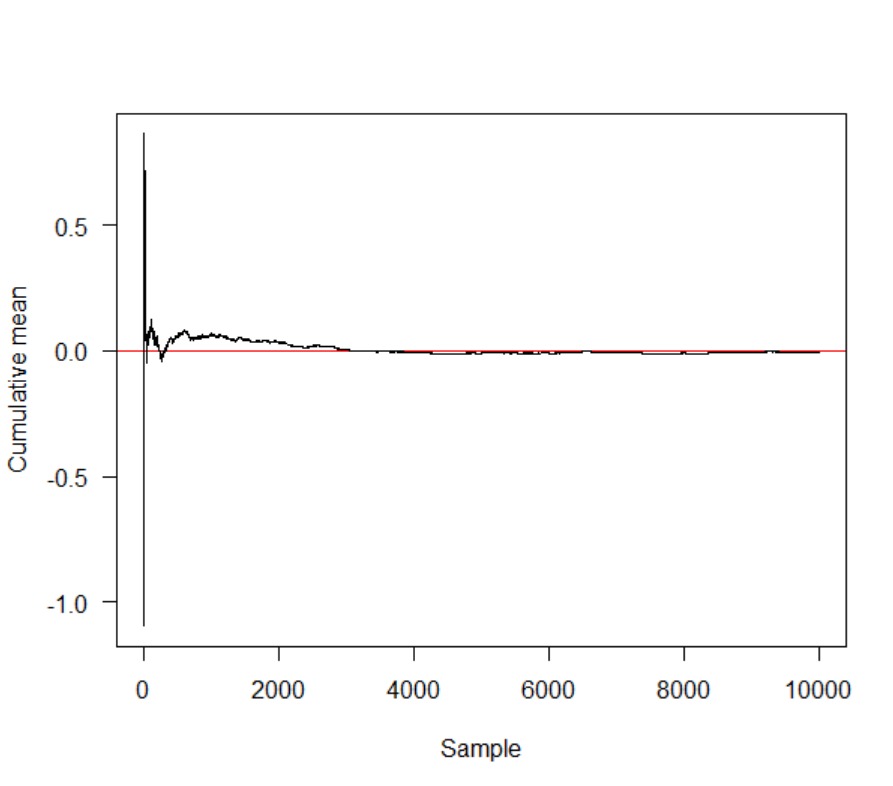
**Inference**

* In statistics, Markov chain Monte Carlo (MCMC) methods comprise a class of algorithms for sampling from a probability distribution.
* By constructing a Markov chain that has the desired distribution as its equilibrium distribution, one can obtain a sample of the desired distribution by recording states from the chain.
* From this we can interpret that one can reduce weight of 1.19 kg in each diet type.

**CUMSUM**



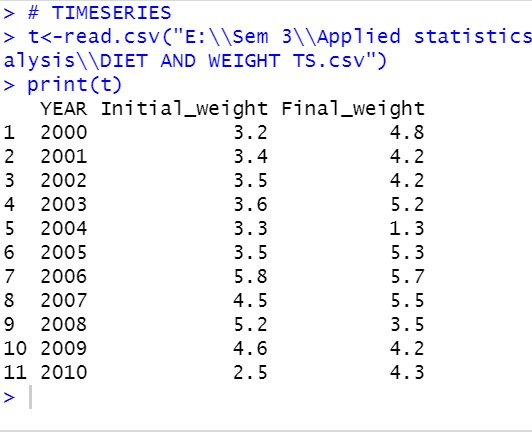




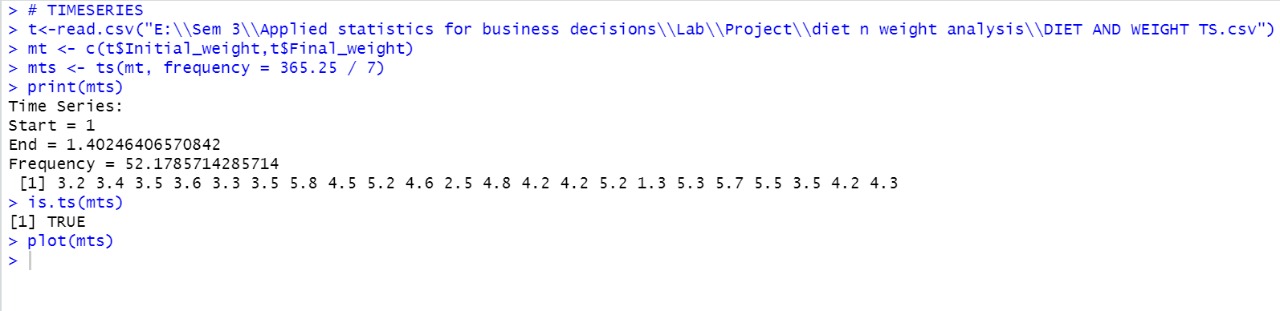
**Inference:**

* If Upward or downward trends develops the process mean has shifted and the process may be affected by special causes.
* From this chart we can interpret in between 0 and 2000 the process will be affected from special causes

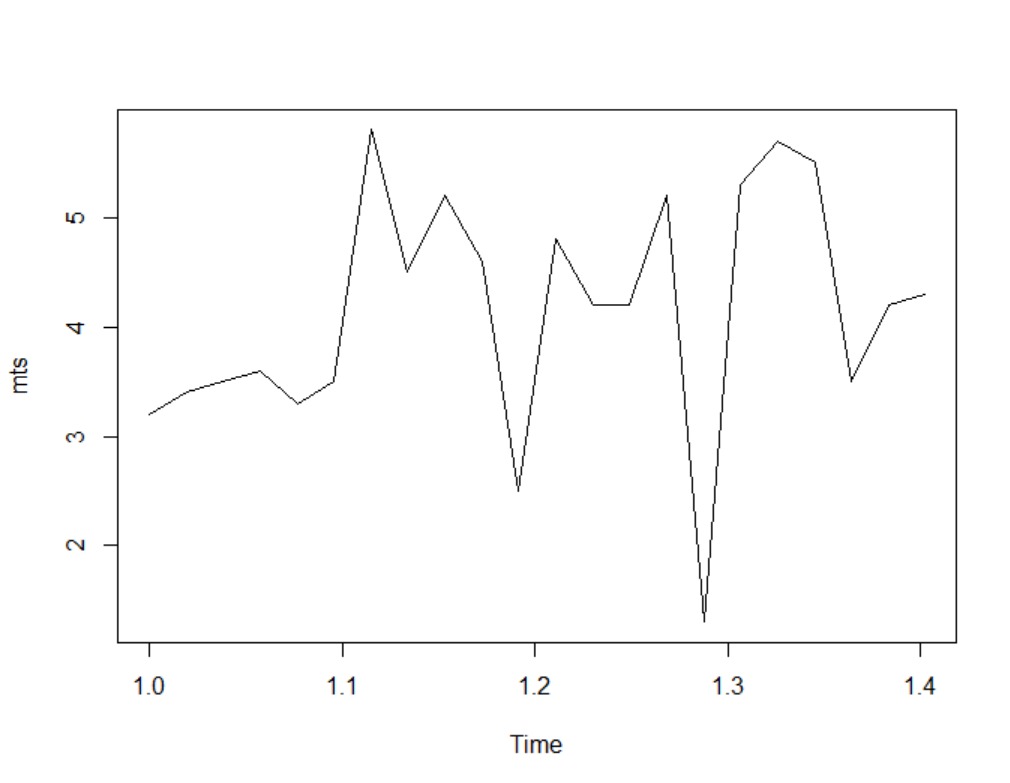
**Time series**



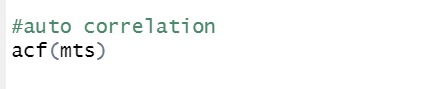
Time series analysis

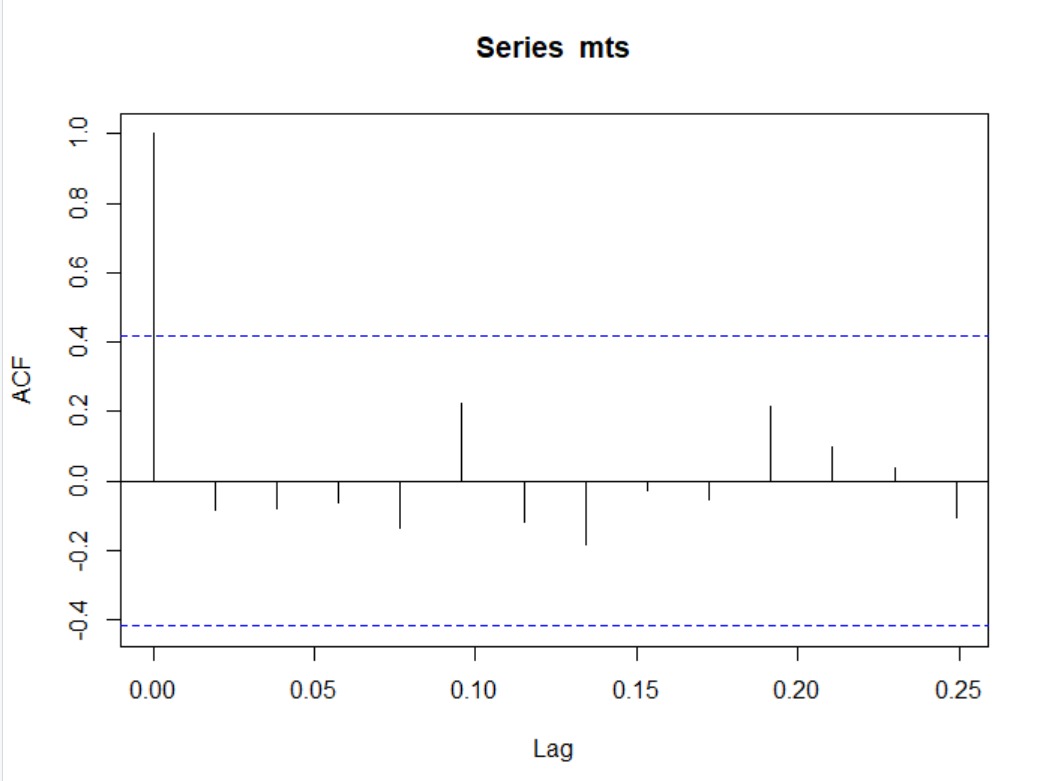


Interpretation



**Auto correlation**

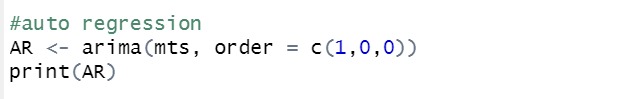


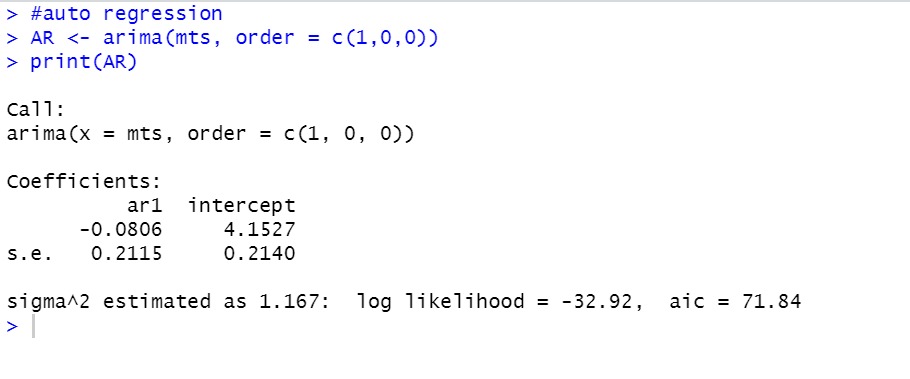


**Inference**

* An autocorrelation plot is designed to show whether the elements of a time series are positively correlated, negatively correlated, or independent of each other.
* Therefore this auto correlation analysis suggests that some diet type may give positive results but there must be some negative outcomes also
* Here in this graph it shows that diet type C has more positive values , So the diet type C is more effective and it is the preferable diet type.

Auto regression





**Inference**

* An autoregressive (AR) model used for forecasting when there is some correlation between values in a time series and the values that precede and succeed them.
* From this we can interpret that type C shows gradual degradation over the years . From this we can infer the Diet type C is more effective for weight loss.

**Result:**

* In this project we can infer that Diet C is very

effective in loosing weight.

* We can interpret that one can reduce weight of 1.19 kg in each diet type.
* People who are practicing diet having weight ranges between 55kg-80kg mostly lies on the height range160cm to 175cm.
* In Diet C one can able to reduce upto 3kg.
* We can infer that height is dependent on weight.
* Diet type A,B,C are independent of each other.

**Conclusion:**

From this, We can conclude that Diet type C is

more effective .So we recommend diet type C to

practice to maintain the body in a healthy state.