**DIET AND WEIGHT LOSS**

**PROBLEM STATEMENT:**

A clinical trial is run to compare weight loss programs and participants are randomly assigned to one of the comparison programs and are counselled on the details of the assigned program. Participants follow the assigned program for 8 weeks. The outcome of interest will be weight loss, defined as the difference in weight measured at the start of the study (baseline) and weight measured at the end of the study (8 weeks), measured in pounds. Three popular weight loss programs are considered. The first is a low calorie diet. The second is a low fat diet and the third is a low carbohydrate diet. After 8 weeks, each patient's weight will be again measured and the difference in weights is computed by subtracting the 8 week weight from the baseline weight. Positive differences indicate weight losses and negative differences indicate weight gains.

A clinical diet is designed to help people to reshape their lifestyle by adopting healthy new habits and breaking unhealthy old ones. The goal is to make simple pleasurable changes that will result in a healthy wait that you can maintain for the rest of your life.The purpose of the Clinic Diet is to help people lose excess weight and to find a way of eating that we can sustain for a lifetime. It focuses on changing daily routine by adding and breaking habits that can make a difference in our weight.

**STATISTICAL METHODS USED:**

Variousstatistical concepts are used to interpret the weight loss and diet . The tools are Generalized Linear Regression Model, ANOVA, Sign test, Kruskal Wallis test, Man-Whitney Wilcoxon test, Spearman’s rank Correlation Coefficient and Wilcoxon signed rank test.The analysis of variance is done by one way and two way anova. They give us the significant difference between the chemicals that influence the taste and quality of wine. The other tests give whether the variables are identical or not. The answers from the tests are used for the correction in the quantity of chemicals of wine.

**ANOVA:** ANOVA or Analysis of Variance is a group of statistical models to test if there exists a significant difference between means. It tests whether the means of various groups are equal or not. In ANOVA, the variance observed in a particular variable is partitioned into different components based on the sources of variation. An important fact to note is that while we use ANOVA to find out whether the means differ significantly. Here we have used one-way and two-way ANOVA. The one-way ANOVA compares the means between the groups you are interested in and determines whether any of those means are statistically significantly different from each other. Specifically, it tests the null hypothesis:

One-way ANOVA Null Hypothesis

Where µ = group mean and k = number of groups. If, however, the one-way ANOVA returns a statistically significant result, we accept the alternative hypothesis (HA), which is that there are at least two group means that are statistically significantly different from each other.The two-way ANOVA compares the mean differences between groups that have been split on two independent variables (called factors). The primary purpose of a two-way ANOVA is to understand if there is an interaction between the two independent variables on the dependent variable. The interaction term in a two-way ANOVA informs you whether the effect of one of your independent variables on the dependent variable is the same for all values of your other independent variable (and vice versa).

**SIMPLE LINEAR REGRESSION:** Simple linear regression is useful for finding relationship between two continuous variables. One is predictor or independent variable and other is response or dependent variable. It looks forstatistical relationship but not deterministic relationship. Relationship between two variables is said to be deterministic if one variable can be accurately expressed by the other.

*Y(Pred) = a + b\*x*

**MAN-WHITNEY WILCOXON TEST:** A popular nonparametric test to compare outcomes between two independent groups is the Mann Whitney U test. The Man Whitney U test, sometimes called the Mann Whitney Wilcoxon Test or the Wilcoxon Rank Sum Test, is used to test whether two samples are likely to derive from the same population (i.e., that the two populations have the same shape). Some investigators interpret this test as comparing the medians between the two populations. Recall that the parametric test compares the means (H0: μ1=μ2) between independent groups. Where H0 states that the two populations are equal and H1, the two populations are not equal.

**SIGN TEST:** The Sign test is a non-parametric test that is used to test whether or not two groups are equally sized. The sign test is used when dependent samples are ordered in pairs, where the bivariate random variables are mutually independent. The sign test is considered a weaker test, because it tests the pair value below or above the median and it does not measure the pair difference.

**KRUSKAL WALLIS TEST:** The Kruskal-Wallis test is a nonparametric (distribution free) test, and is used when the assumptions of one-way ANOVA are not met.  Both the Kruskal-Wallis test and one-way ANOVA assess for significant differences on a continuous dependent variable by a categorical independent variable (with two or more groups). Kruskal-Wallis test can be used for both continuous and ordinal-level dependent variables.

**WILCOXON SIGNED RANK TEST:** The Wilcoxon Sign test is a repeated measures test of dependency. This test is mathematically similar to conducting a Mann-Whitney U-test. It is a non-parametric analysis that statistically compared of the average of two dependent samples and assess for significant differences and also a non-parametric alternative of the dependent samples t-test.

**SPEARMAN’S RANK CORRELATION COEFFICIENT:** The Spearman’s Rank Correlation Coefficient is the non-parametric statistical measure used to study the strength of association between the two ranked.

**HYPOTHESIS:**

Hypothesis refers to the claim or assumption or made by analysts to prove if their point of view on the analysis is right or wrong.

For the Diet and Weight Loss Analysis, the following assumptions are made:

* Null hypothesis: The weight loss caused by the diets are same irrespective of the diet types a person is undergoing.
* Alternative hypothesis: The weight loss caused by the diets are different for different diet types.

**DATA:**

Data set is collected from the website : <https://www.sheffield.ac.uk/polopoly_fs/1.570199!/file/stcp-Rdataset-Diet.csv>

This data set has the right attributes for diet and weight loss analysis. The main attributes such as initial weight, final weight, and diet type are present in this dataset. Furtherly, other attributes such as gender, height and age play significant roles here. In this project, the dataset consists of observations made on 78 persons undergoing 3 types of diets like Diet A, Diet B, Diet C.

**STATISTICAL MODEL:**

Statisticalmodels are used by analysts to make the general public to understand their analysis effectively.

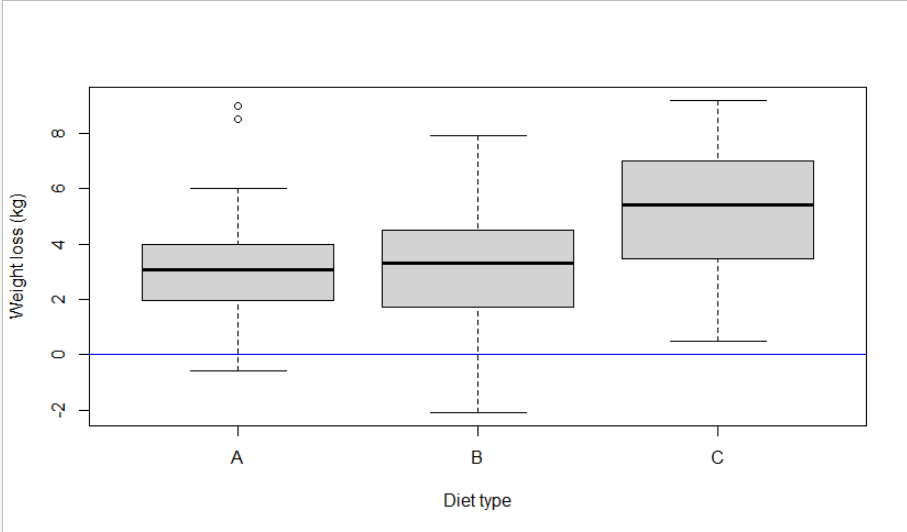


FIGURE 1

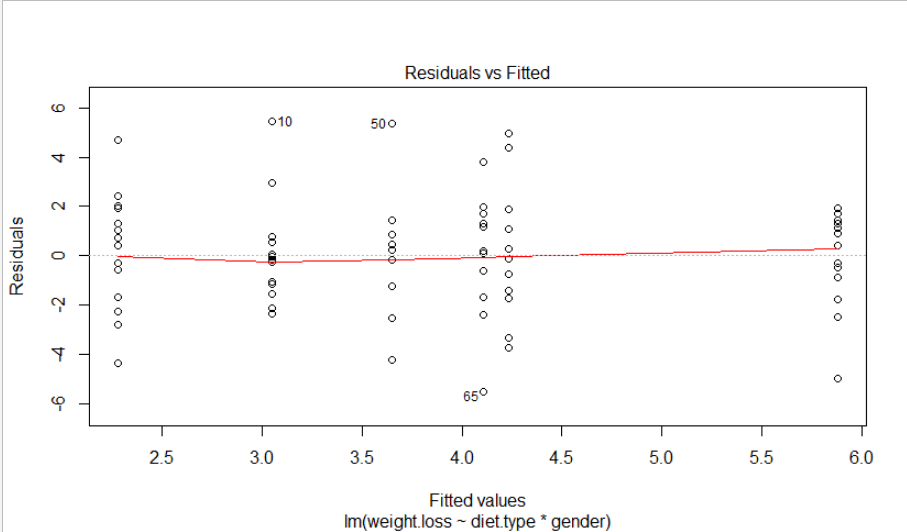


FIGURE 2

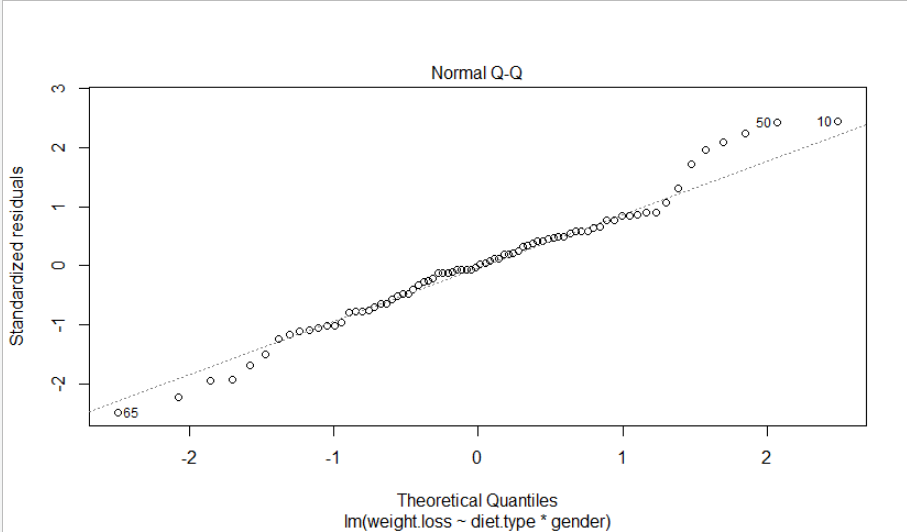


FIGURE 3

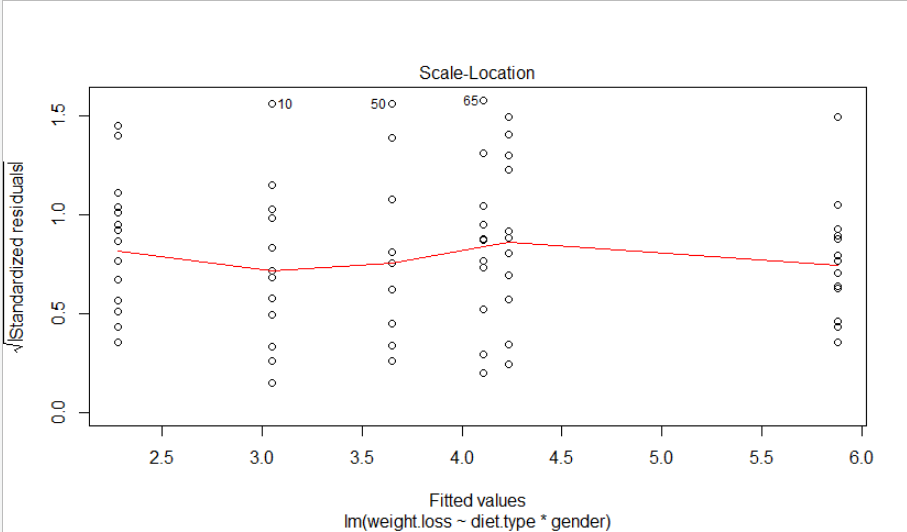


FIGURE 4

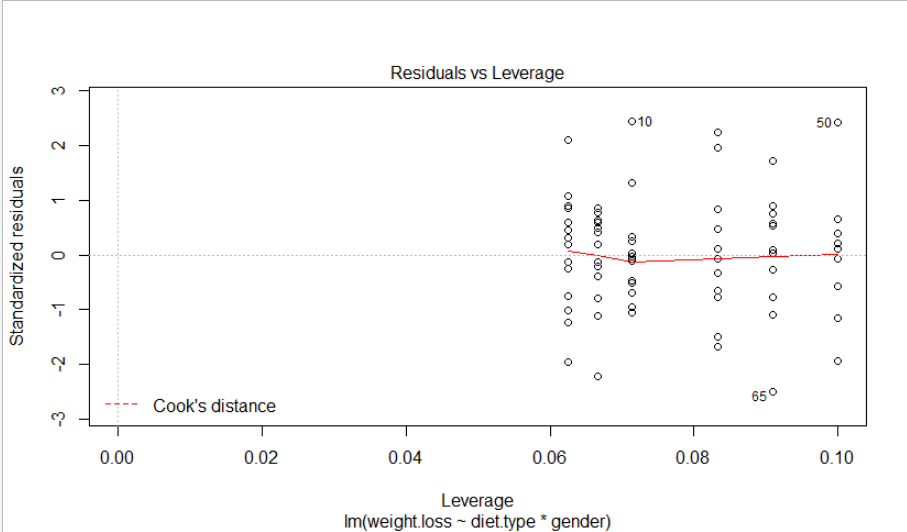


FIGURE 5

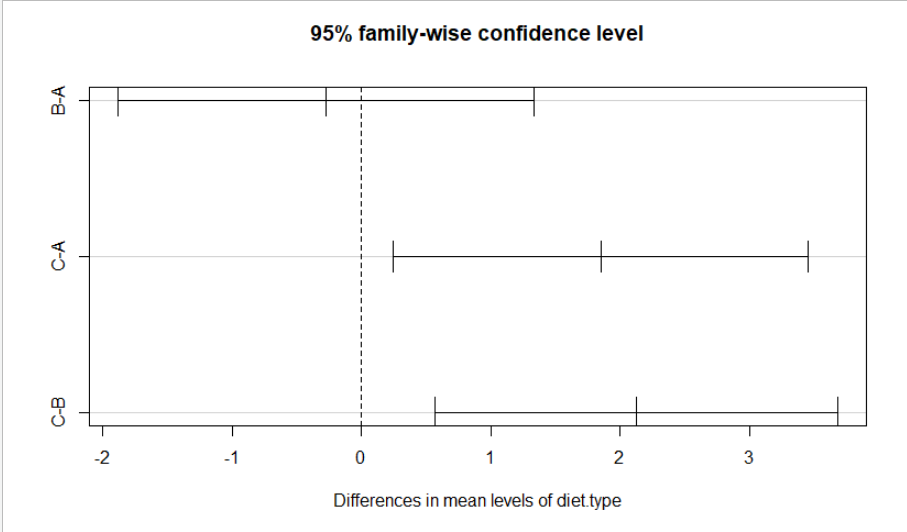


FIGURE 6

* FIGURE 1 represents Box plot
* FIGURE 2,4 and 5 represents Linear Models
* FIGURE 3 represents Normal Q-Q plot
* FIGURE 6 represents Tukey Honest Significant Differences

**RESULTS:**

From the results obtained from the code, the null hypothesis get rejected (i.e.), “All diets yield same results” is proved to be false here. Hence the alternative hypothesis stating that “Different diets have different effects on people” is accepted here. It is inferred that of all the diets, Diet B gives the maximum weight loss compared to Diet A and Diet C. This interpretation is made based on the results obtained from ANOVA (one-way) and the subsequent tests made.

**CONCLUSION:**

From this project, the conclusion made is that Diets of different types have different impacts on people undergoing them and that the diets are independent of one another.