Weight Loss in Relation to Three Different Diets

I. Introduction

Every year, going to the gym and getting in shape is one of the things that's on people's new year's resolutions. Usually when people go to the gym, one of their goals is to lose weight. Now even though going to the gym does contribute to weight loss, the diet that the individual chooses is also important in losing weight. The data collected in this experiment has 76 participants that went on one of three different diets. The time period of this experiment would conclude after six months from the participants going on their diets. During this experiment, the response variable is the difference in pounds of how much weight the participants gain or loss. The explanatory variable is the three different diets. Factors that's also being considered but not shown is if these participants have any medical issues because sometimes these factors can affect how much weight the participants gain and lose compared to a healthy person with no problems. But in this test we assume that the participants all are healthy while taking one of the three diets. The question we are trying to solve is whether or not having different diets would result in the same weight loss. This question is interesting because people have speculated that certain diets lead to better results in losing weight. If the data from this experiment can prove that there isn't much of a difference of weight loss for any diets, people would be relieved that any diet they choose can lose around the same amount of weight as other diets. For this experiment, we are using ANOVA because the 76 participants are randomly selected and assigned to one of three diets that are independent from one another. We will also be interpreting negative weight loss as weight gains

Obviously if there are outliers provided in the data, we would be determined that those outliers have "factors" that make the individual different from the rest. Randomly selected individuals being selected for data sampling isn't always going to be perfect because each individual is different health wise. To better understand the experiment at hand, we will determine that we will be using a 95% confidence interval. We are also using an alpha level of 0.05 to test whether diet type has an effect on weight loss. Here are statistics of what we found in the dataset.

II. Summary of Data

We are summarizing the data because it shows us the mean of weight loss for the three diets, the standard deviation for how dispersed the data is for each diet, and the sample of how many participants were for each diet.

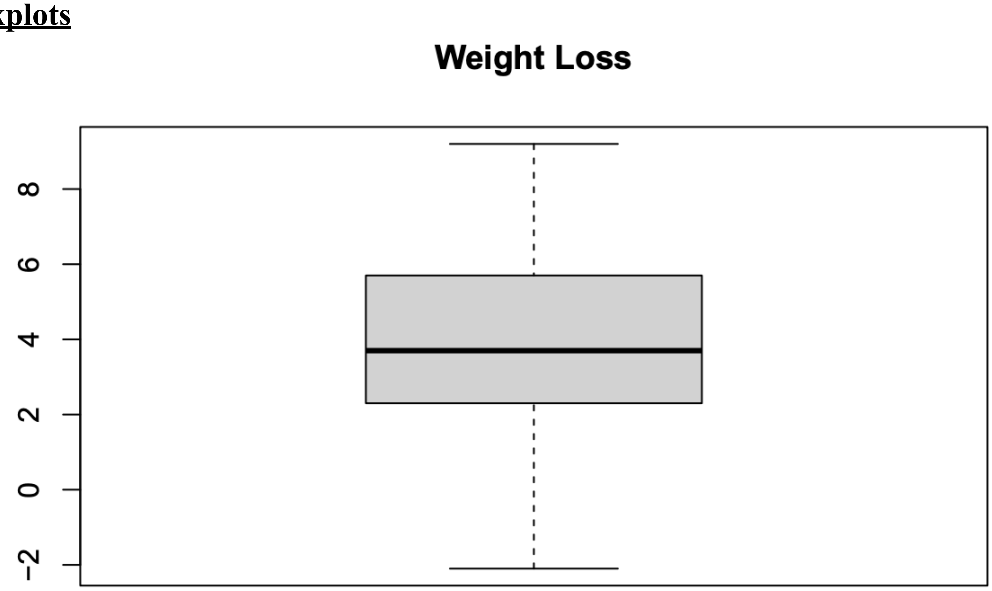
Graphs are insufficient because they are used for easy visualization. Having a summary of the data collected help gives us an additional in-depth look as to what is shown visually. Although graphs such as histogram, box plot graphs, or normal QQ plot will give us quantitative evidence, the interpretation of these graphs visually will allow us to look at the center (median/mean) and spread of our data set, and get a sense of the variability within the data.

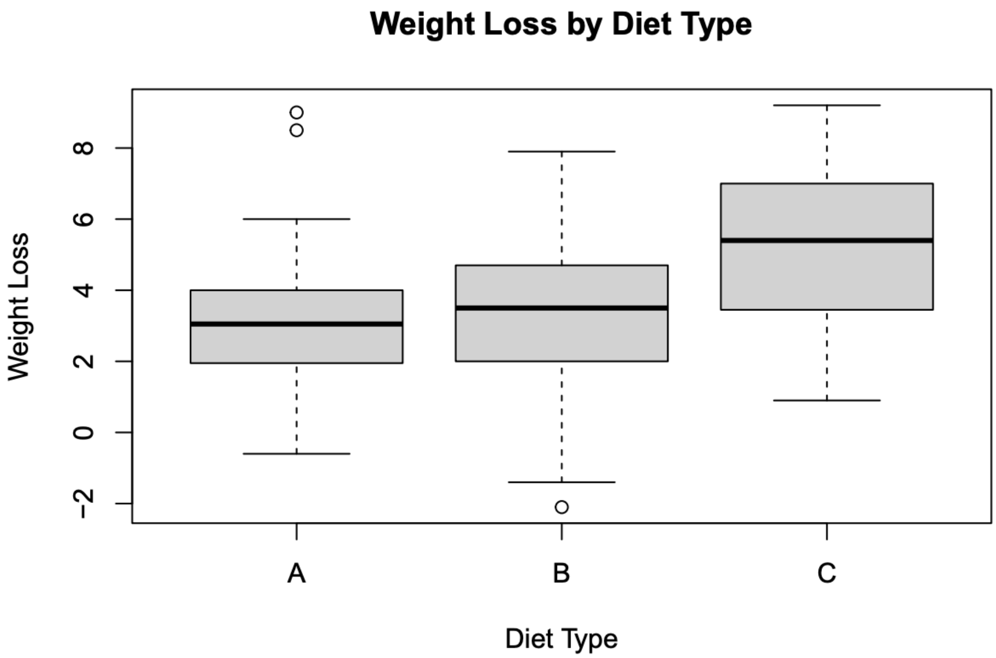
II. 1. Summary Statistics

Summary Statistics of Weight Loss by Diets

	A	B	C	Overall
Sample Mean	3.3000	3.2680	5.2333	3.9763
Sample Std. Dev.	2.2401	2.4645	2.2477	2.4732
Sample Size	24	25	27	76

II. 2. Boxplots



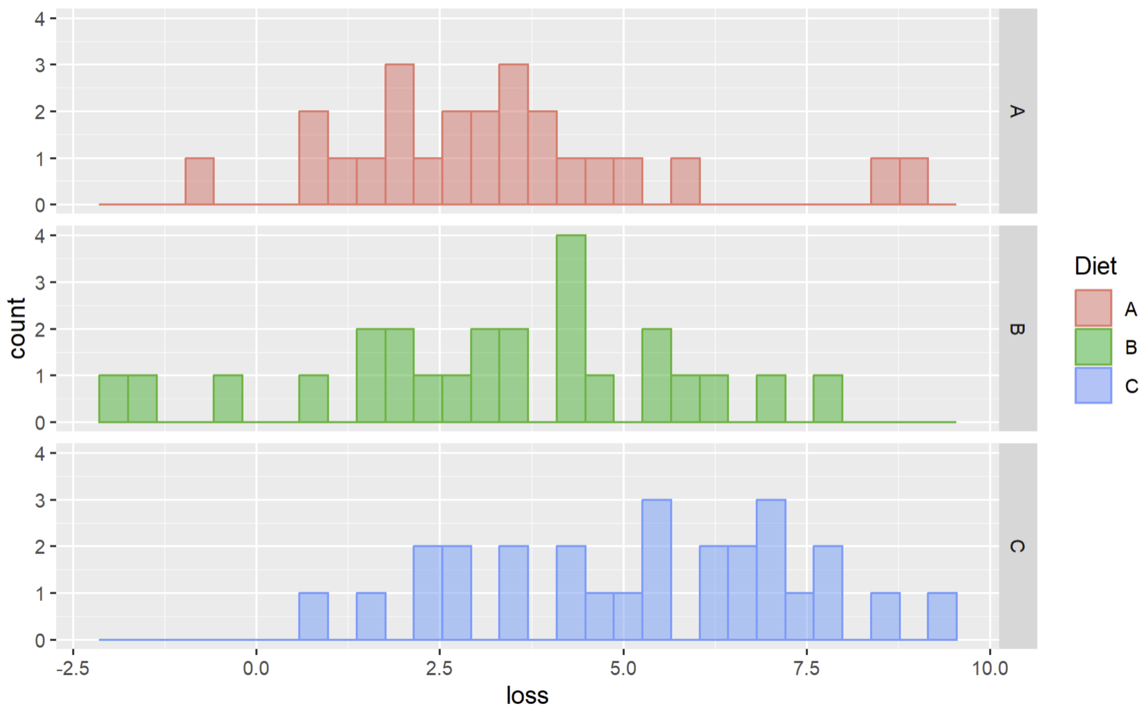


From the boxplots, the first thing that we noticed is the outliers in Diet A and Diet B. In Diet A, there seem to have 2 outliers higher than the upper limit line with weight loss recorded higher than 8 pounds. In Diet B, there seems to be one outlier smaller than the lower limit -1 pounds of weight loss. Diet C does not seem to contain any outliers. In terms of range of the data, Diet B has the largest range yet smallest minimum value (-1 as minimum and 8 as maximum). Diet C seems to have a slightly smaller range and highest maximum value (1 as minimum and 10 as maximum). Diet A has the smallest range. In terms of interquartile range, which refers to the distance between 75th percentile and 25th percentile, Diet C has the largest interquartile range, followed by Diet B and then finally Diet C has the smallest interquartile range. We can see from the grouped box plot that the median weight loss for diet C is significantly the largest among 3 groups, followed by Diet B and then Diet A. The interquartile range for diet C is larger than that of both diet A and B, and the interquartile range for diet A is the smallest out of the diets. From observing this box-and-whisker plot, it seems like participants opting in for Diet C might result in the highest weight loss and thus better result in controlling weight, as compared to Diet B and Diet C. However, further analysis in single factor ANOVA is needed to determine the variance within groups and between group means and overall means.

II. 3. Overall Histogram

The histogram represents the randomly sampled and observed diet type of all 76 participants, by the average weight loss they occur. Here each bar represents a range of 1, meaning 1 pound of weight loss. Therefore, a negative weight loss here would represent a weight gain in pounds (e.g. -2 weight loss equates to 2 pounds increase in weight). The bar height represents the frequency of participants exhibiting that particular weight loss in pounds. Based on our observation of the histogram, we are able to identify that the graph is clustered towards the middle with a slight skew to the left. More specifically, most of the weight loss is clustered towards 1 to 5 pounds weight loss. Therefore, it is safe for us to assume that the overall mean is somewhere between 1 to 5 pounds range as well. We can see that on the left tail of the graph, there are 1 observation with weight loss between -4 and -2, and 3 observations with weight loss between -2 and 0. These couple observations might potentially be outliers. We will investigate this further and eliminate outliers in the next part when we expand the overall histogram into histograms by groups.

II. 4. Histograms by Group

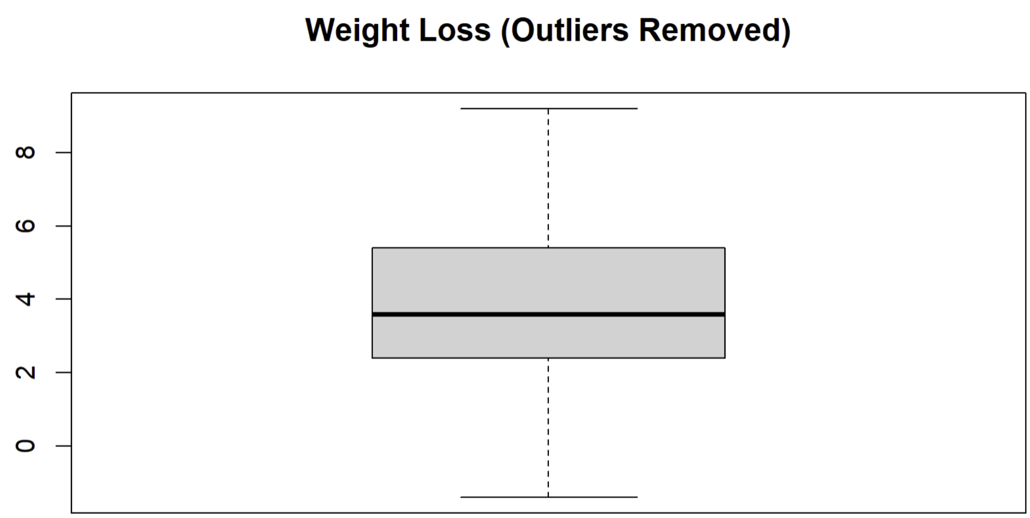


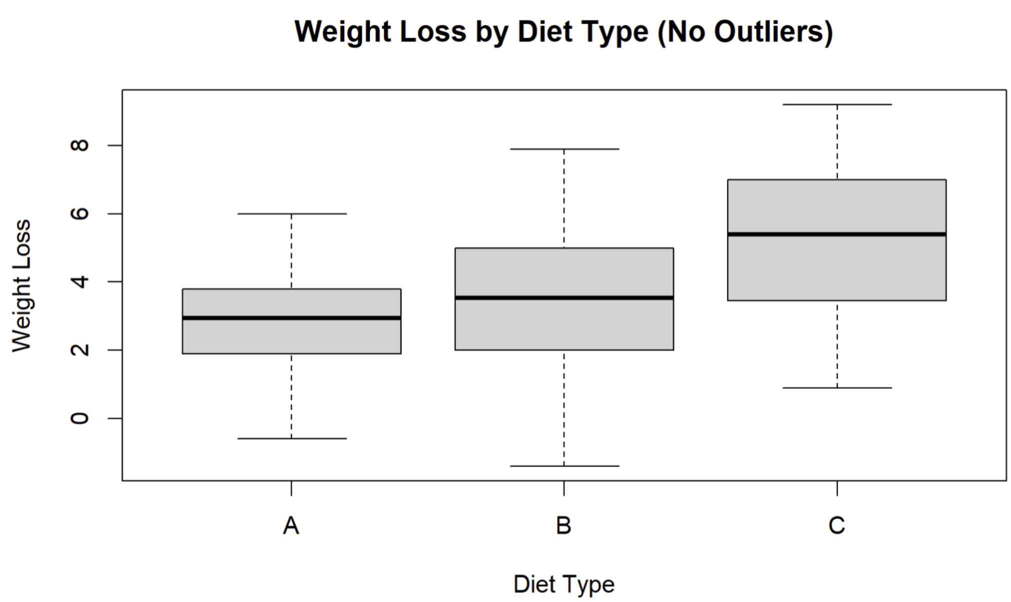
These three graphs show the weight loss of the participants based on the three different diets they were placed on, namely Diet A, Diet B, Diet C. All 3 graphs have bars with widths that represent a range in weight loss that represent 0.1 pounds. Diet A seemed to be slightly skewed to the right. Diet B is skewed to the left. Diet C seems to be slightly skewed to the right. The median weight loss for diet C is larger than that of diets A and B. However, there is not as much of a difference for the median between diets A and B. The interquartile range for diet B is larger than that of the other two groups. The largest range of weight loss is group with diet C. The smallest range of weight loss is group with diet B. Diet A seems to have some outliers to the right while Diet B seems to have some outliers to the left.

III. Diagnostics

III. 1. Removing Outliers

We determined that our dataset had outliers and we decided to remove the outliers. There were 3 outliers.





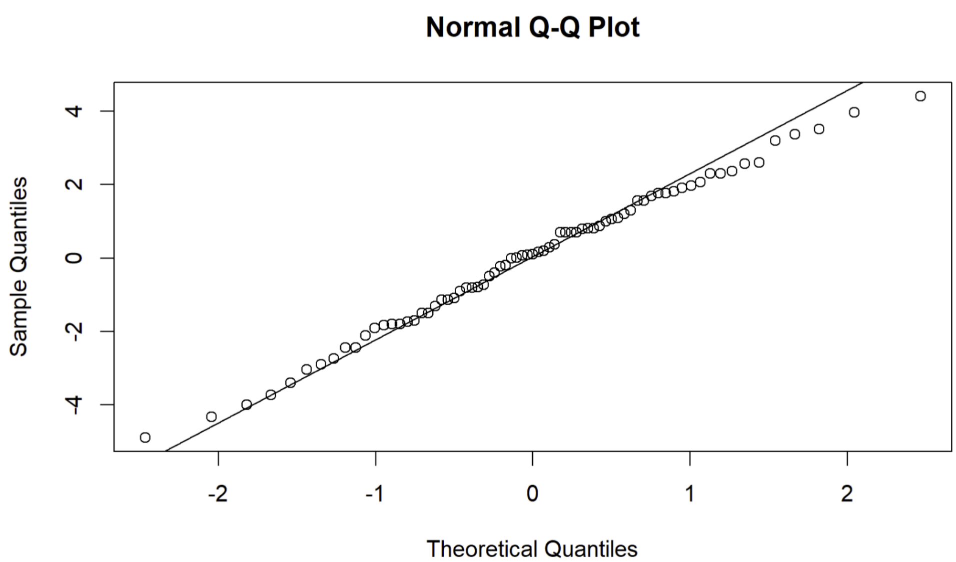
III. 2. Summary Statistics Without Outliers

Summary Statistics of Weight Loss by Diets (Outliers Removed)

	A	B	C	Overall
Sample Mean	2.8045	3.4917	5.2333	3.9288
Sample Std. Dev.	1.5506	2.2434	2.2477	2.2851
Sample Size	22	24	27	73

From this point forward we will be using and referring to the dataset without outliers.

III. 3. Normal QQ Plot (Normality)



The QQ Plot helps to provide a visual representation of whether or not our data follows the normal distribution. What we are essentially looking for here is whether or not the points on the plot follow the straight line. On one hand, if they do fit onto a straight line nicely then we can safely assume that our data is likely to be normal. On the other hand, if the points follow a random pattern then we can deduce that there are outliers coming from the data set. This QQ plot has most of the points fairly close to the line with a few stray points and the tails of the QQ plot seem to slightly deviate from the normal line. We can assume an approximately normal distribution.

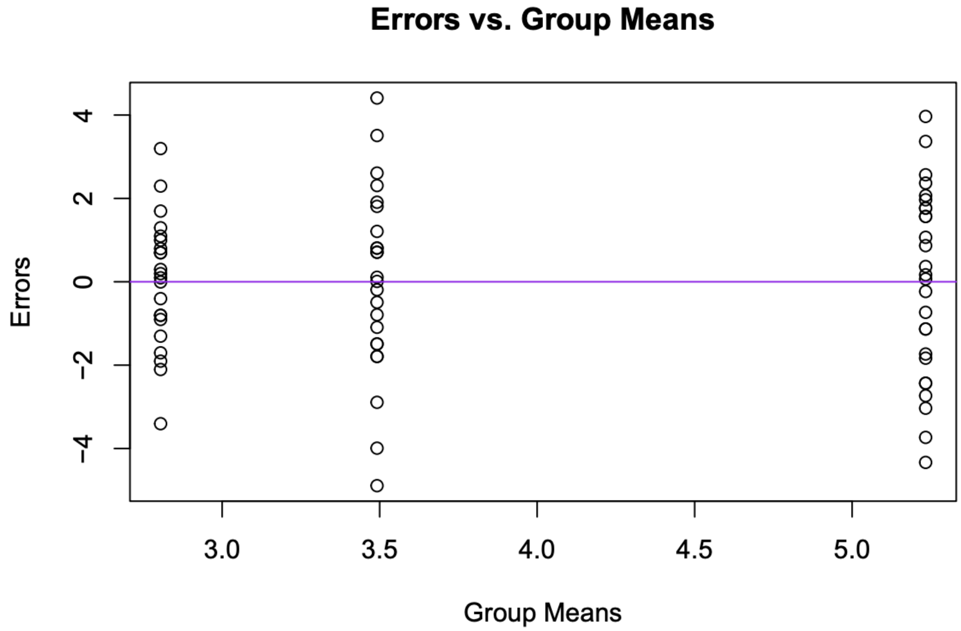
III. 4. Shapiro Wilks Test (Normal Distribution of Errors)

H_0: The errors are normally distributed.
H_a: The errors are not normally distributed.

From performing the Shapiro Wilks Test, we get a p-value of 0.909 and W = 0.99149.

Since W is really close to 1 and this p-value is larger than our alpha, we cannot reject the null hypothesis. Therefore, we can conclude that our errors are normally distributed.

III. 5. Constant Variance Plot



In ANOVA, the fitted values are equal to the means for that group. We plotted the errors vs. the group means to assess if there is equal variance. From the plot, it suggests roughly equal spread and constant variance.

III. 6. Brown Forsythe Test (Constant Population Variance)

H_0: The population variances by group are equal.
H_a: At least one population variance is not equal.

From performing the Brown Forsythe Test, we got a p-value of 0.1300887. Since this p-value is larger than our alpha, we cannot reject the null hypothesis. Therefore, we can conclude that the population variances are equal.

Summary:

The analysis of the dataset, which focuses on the weight loss outcomes of 73 individuals assigned to one of three different diets (labeled A, B, and C), aims to determine the most effective diet for weight loss. The subjects' weight loss are independent from one another and the subjects are randomly chosen. To assess the normality of the weight loss data within each diet group, a QQ plot was utilized. This plot visually examines whether the data follows a normal distribution, with points ideally aligning along a straight line. The QQ plot indicated that while most points closely followed the line, there were a few outliers, and the tails of the plot deviated slightly from the normal line. Subsequently, the Shapiro-Wilks Test was performed, yielding a p-value of 0.909 and a W statistic of 0.99149, suggesting that the errors are normally distributed. Furthermore, ANOVA was conducted, and the fitted values, which represent the means for each diet group, were plotted against the errors to assess for equal variance. The plot exhibited roughly equal spread and constant variance across groups. Additionally, the Brown-Forsythe Test yielded a p-value of 0.1300887, indicating that the population variances are equal. Therefore, despite minor deviations observed in the QQ plot, the normality assumption was met, and the homogeneity of variances assumption was upheld, allowing for reliable inference regarding the effectiveness of the three diets for weight loss.

IV. Analysis

IV. 1. ANOVA Table

	df	Sum Sq	Mean Sq	F Value	Pr(>F)
Diet	2	78.342	39.171	9.2133	0.0002806
Residual	70	297.608	4.252		
Total	72	375.95	5.221528		

IV. 2. Null and Alternative Hypothesis Test

Our goal for this study is to determine whether or not there is a difference in average population weight loss in pounds between three diets. Therefore our null and alternative hypotheses are:

H_0: The average population weight loss is the same across Diet A, Diet B and Diet C.
H_a: At least one of the diets (A, B and/or C) has a different average population weight loss.

From our ANOVA Table, we have a p-value of 0.0002806. This p-value is less than our alpha of 0.05. At the 5% significance level, we at least know that 5% of our overall sample, which is at least three samples, has a drastically different weight gain or loss than the rest.

IV. 3. Power

$$\phi = \frac{1}{\sigma_\epsilon}\sqrt{\frac{\sum_i n_i(\mu_i - \mu_.)^2}{a}}$$

This formula gives us the non-centrality parameter where we use our sample values to estimate the assumed true standard deviations of the errors, the overall population mean for group i, and the overall population means of all Y_{ij}. Using those values, the approximate power is 0.971908.

d.f. {num} = a-1 = 2
d.f. {denom} = nT - a = 76 - 3 = 73

IV. 4. Confidence Interval

$$(\overline{y}_a - \overline{y}_b) \pm t_{\alpha/(2k)}\sqrt{MSE(1/n_a + 1/n_b}$$

This formula is used to determine the range of how much of a difference each group means from each other.

All pairwise confidence intervals:

Pair	Confidence Interval
μ_A-μ_B	(-1.900431, 0.5260309)
μ_A-μ_C	(-3.609355, -1.248245)
μ_B-μ_C	(-2.894732, -0.5884679)

V. Interpretation

V. 1. P-Value Test

If the average population weight loss is the same across Diet A, Diet B and Diet C, we would observe our data or more extreme with probability < 0.0001. Since our p-value is less than our alpha level o $\alpha = 0.05$, we reject the null hypothesis and conclude that there's statistical evidence that at least one diet has a different average population weight loss.

V. 2. Confidence Intervals

We are 95% confident that the true average difference in weight loss in pounds for diet A compared to diet B is between -1.900431 and 0.5260309 pounds. This means that the subjects who had diet A either lost weight, gained weight, or did not have a change in weight compared to those who had diet B.

We are 95% confident that the true average difference in weight loss in pounds for diet A is less than diet C by between -3.609355 and -1.248245 pounds. This means that the subjects in diet A lost less weight than those in diet C.

We are 95% confident that the true average difference in weight loss in pounds for diet B is less than diet C by between -2.894732 and -0.5884679 pounds. The subjects in diet B lost less weight than those in diet C.

V. 3. Power

Power is the probability to correctly reject the null hypothesis when it is false. Our power value is 0.971908. This means that there is about a 97.19% chance that we correctly rejected the null hypothesis. This supports our decision.

VI. Conclusion

This study offers an in-depth analysis of how various diets affect weight loss, drawing on data from 73 individuals following one of three distinct dietary treatments: A, B, and C. The conclusion provides a comprehensive understanding of the impact of different diets on weight loss, revealing significant insights into the effectiveness of these dietary plans. The statistical analysis includes the ANOVA test, the confidence intervals for the differences in mean weight loss between the diets.

The ANOVA test, focused on figuring if there are any diets would result in a significantly different average population weight loss compared to the others. Our results indicated that at least one of the diets resulted in a different average population weight loss compared to the others with a p-value of 0.0002806, being considerably lower than the alpha level of 0.05, leading to the rejection of the null hypothesis.

Further, the confidence intervals for the differences in mean weight loss between the diets provide a statistical prediction for how large the average difference in mean weight loss for different diets will be. Specifically, the confidence intervals indicate that Diet C was more effective in promoting weight loss compared to Diets A and B. This is supported by the fact that the confidence intervals for the differences in average weight loss between Diet C and the other two diets do not include zero and are skewed towards Diet C being more effective.

In summary, this study provides statistically significant evidence that different diets can lead to varying levels of weight loss. While individual results may vary and are influenced by numerous factors including genetics, lifestyle, and adherence to the diet, our analysis suggests that Diet C may be more effective for weight loss compared to Diets A and B. This finding could be valuable for individuals seeking to lose weight, as well as for dietitians and health professionals advising on weight loss strategies.