*An Analysis of the Significance of Walking based on Diet Types*

*A\_group 33*

*Simpson Chiwashira, Trishan Banerjee, Affan Idowu Kareem, Olawuyi Pelumi Emmanuel, Jennifer Oge Kela-Onwuchekwa*

Friday, 13 Jan 2023

# Abstract

Walking is often neglected as an effective form of exercise in today’s world that greatly promotes a healthy lifestyle. In this study, we wanted to analyze the significance of the number of steps a person took as a form of exercise based on their dietary and fitness goals. This research asks if there is a difference in mean between the number of steps a person takes based on their dietary type. We analyzed the fitness data of a healthy male for 130 days. The dataset was taken from data. world. Although there seemed to be a clear difference in the number of steps for the different diet types, a pairwise comparison test revealed a significant difference in only two of the diet types, i.e., between Reverse dieting and Cutting. The study proves the significance of walking as an effective method of exercise that can be scaled depending on a person’s goals.

Table of Contents

[Abstract 1](#_Toc124512539)

[Introduction 3](#_Toc124512540)

[Visualization 5](#_Toc124512541)

[Analysis 7](#_Toc124512542)

[Conclusion 9](#_Toc124512543)

[Limitations and Further work 9](#_Toc124512544)

[References 10](#_Toc124512545)

[Citations 10](#_Toc124512546)

[Figures 10](#_Toc124512547)

[Tables 10](#_Toc124512548)

# Introduction

Walking is a simple, free and effective form of cardiovascular exercise. Walking is an often-overlooked form of exercise. According to the NHS, brisk walking can help build stamina, burn calories and improve heart health (Walking for health, 2022).

The purpose of this study was to find the difference in the number of steps an average male must take according to their specific form of diet to reach their fitness goals. The fitness goals have been categorized as weight gain, weight loss, and maintaining current weight. For these, the specific forms of diet are reverse dieting (for gaining weight), cutting (for losing weight), and maintaining (to maintain the same weight). Health and fitness are very popular topics, and proving the importance of walking, an underrated form of exercise makes this study interesting.

Our Research question for the study is as follows:

“Is there a difference in mean between the number of steps taken by an individual and their diet type depending on their fitness goals?”

The Null Hypothesis states,

“There is no difference in the mean between the number of steps taken by an individual and their diet type depending on their fitness goals.”

The Alternate Hypothesis states,

“There is a difference in the mean between the number of steps taken by an individual and their diet type depending on their fitness goals.”

Our dataset is from Data.world. The URL for our dataset is “<https://data.world/selcukyilmaz/fitness-dataset-of-healthy-man-130-days>”. The dataset's goal is to observe a healthy man's data for 130 days. The dataset has 21 columns that categorize the caloric intake, the workouts performed, the water intake, the macronutrient breakdown, the sleep time, and other such data of a man daily for 130 days. Among these values, we chose to study the number of steps taken, our dependent variable, an interval datatype, and the corresponding diet type the man was following, and our independent variable, a nominal datatype. Our dataset was found to be normally distributed as mentioned in the following sections of this report.

We used the pairwise t-test to analyze our dataset. The results show that there is a significant difference in the mean steps between the Cutting and Reverse Dieting forms of diet but there was so significant difference between Cutting and Maintaining. Thus, if the individual was looking to burn more calories, they would increase their daily step count, and vice versa if they were trying to burn fewer calories. This proves that depending on an individual’s diet and fitness goals, they can incorporate walking into their exercise regime as it has proven to have a significant impact.

This paper has 4 different sections. The first section is the “*Visualization”* section which describes our data and visualizes it in a graphical form. The next section is the “*Analysis”* section which highlights the methods used in this study and outlines the results obtained from our methods. This is followed by the “*Conclusion*” section which details our conclusions drawn from this study. The final section is “*Limitations and Further Work*” which details the limitations of our study and the scope of future works.

# Visualization

This section visualizes our dataset in the form of box plots. The dataset used for this study comprises the fitness data of a healthy man obtained for 130 days. The data has a total of 21 columns and 133 rows of data (including headings). The two columns that we used for the visualizations are diet type and the number of steps.

The data is used to investigate the mean between the number of steps taken by an individual and their diet type depending on their fitness goals. The data markers are identified as “Cutting”, “Maintain” and “Reverse Dieting” and are represented in a box plot as shown in Figure 1.  Here, Cutting refers to the process of reducing body weight, Maintain refers to the process of maintaining the same body weight, and Reverse Dieting refers to increasing body weight.

Chart, box and whisker chart

Description automatically generated

Figure 1:Box plot showing the steps assigned based on each diet type.

Chart, histogram

Description automatically generated

Figure 2:Cumulative distribution function (CDF) of steps assigned by all markers.

Figure 2 shows the cumulative distribution function of steps assigned by all markers. The Histogram shows that the steps are skewed maximum between 15,000 and 17,000 steps, which is the usual bell curve of a normal distribution.

# Analysis

We have our visualizations plotted out and we can see some differences. However, we are not convinced and so we have made further analysis. We will start by analyzing the summary of our independent variables and dependent variables. The data is shown below:

|  |
| --- |
| Maintain Cutting Reverse Dieting  Min. :15936 Min. :11267 Min. :10062  1st Qu.:16516 1st Qu.:16284 1st Qu.:14688  Median :17409 Median :18070 Median :15370  Mean :17782 Mean :17558 Mean :15518  3rd Qu.:18501 3rd Qu.:19405 3rd Qu.:16579  Max. :22246 Max. :22900 Max. :19379  NA's :104 NA's :99 NA's :61 |

Table 1: Summary of our independent variable showing all the categories.

We are investigating a comparison of means and the summary of the data seems to suggest a significant difference between reverse dieting and the other two variables. However, maintaining and cutting seem to show a small difference in the mean. So, to investigate further, we will use the appropriate test static provided by the R programming language. Figure 2 shows the histogram, and the distribution curve shows a bell curve-shaped figure, suggesting that the data is normal. Using the Shapiro-Wilk test for normality provided by the R programming language, supported our claim (Shapiro.test: Shapiro-Wilk Normality Test). Table 2 shows the result of the Shapiro test.

|  |
| --- |
| Shapiro-Wilk normality test  data: steps  W = 0.98574, p-value = 0.1855 |

Table 2: Results after running the Shapiro Test for normality of the data.

The p-value of this result is not less than 0.05. and the current p-value is greater than the threshold alpha value of 0.05 suggesting that the data is normally distributed. The appropriate test static for normal parametric data is the student’s test. Our independent variable has three categories hence we cannot use the t.test but instead use the pairwise.t.test.

The pairwise t-test method usually is used to find the comparison of the means of two values. Since we have more than two values, it will check every possible set and computes the p values.

|  |
| --- |
| Pairwise comparisons using t-tests with pooled SD  data: steps and diet.type  Cutting Maintain  Maintain 0.67 -  Reverse Dieting 1.5e-05 8.5e-06  P value adjustment method: holm |

Table 3: P-values resulting from the pairwise t-test with the holm adjustment comparing the difference of means in the steps taken by an individual.

The above output shows the result after performing the pairwise t-test and using the holm adjustment. The table shows the p values of each set of values that have been compared. It further supports our claims that we found in the summary of our dataset as we see that Cutting and Maintaining have a p-value greater than our alpha value (0.05). And with this, there is no significant difference between Cutting and Maintaining. However, there is a significant difference between the other two sets of values, Reverse Dieting vs Maintain and Reverse Dieting and Cutting since the pairing values are less than the alpha value (0.05). The probability of the difference being random is very small. We consider rejecting the null hypothesis in the case of the two significant values (p-value < 0.05) and not in the case where the (p-value > 0.05). As such, there is no general hypothesis that we can consider because of the two cases where we fail to reject the null hypothesis and where we rejected it.

# Conclusion

The results from our analysis suggest that the diet type “Reverse Dieting” which is simply gaining more weight, means that the individual will exercise less. He does not wish to burn any calories by walking more. But when he is trying to cut down his weight and maintain it, he tends to walk more steps. We can visualize this in Figure 2, where we can see that he walks as much as 20,000 steps and on average he walks at least 17,000 steps. The statistics are showing that there is no significance between the diet types in question.

In any circumstance, if we reject the null hypothesis, we can come to an understanding that the relationship is not coincidental for the two cases where “Maintain” vs “Reverse Dieting” and “Cutting” and “Reverse Dieting”. The probability values after the computation in the R programming language show significance and that tells us that it was not a fluke that we came up with the values. However, in the case of “Maintain” and “Cutting”, we can see that the relationship was purely coincidental since the probability is failing to reject the null hypothesis.

When we fail to reject the null hypothesis, the answer is “no” to our research question. That is only true in one case where the significance of the probability value was less than the alpha value. But in the other two cases, it suggests that the data is not enough for us to come up with a comprehensive different conclusion.

# Limitations and Further work

Our study would be more comprehensive if the dataset accounted for more than one individual and a variety of people, this could lead to more promising research in the future.

Further, different adjustment methods can be used for our p-value and analyze the values varying in terms of significance. As seen in the analysis we used the “holm” adjustment and if we use the “Bonferroni” the value changes. Given more time we would try all the methods and see the difference and research more on how they adjust and come to choose the best method.

Further tests can be conducted using the ANOVA test which could help in further analysis to see what the generation of the p-value is. In the future, different tests can be performed to come up with different results.

# References

## Citations

J.P. Royston. (1982). Algorithm AS 181: The W test for Normality. *Applied Statistics*, 176.

J.P. Royston. (1982). An Extension of Shapiro and Wilk's W test for Normality to Large Samples. *Applied Statistics*, 115.

J.P. Royston. (1995). Remark AS R94: A Remark on Algorithm AS 181: The W-test for Normality. *Applied Statistics*, 547.

*Shapiro. test: Shapiro-Wilk Normality Test.* (n.d.). Retrieved from the documentation: https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/shapiro.test

*Walking for health*. (2022). Retrieved January 11, 2023, from https://www.nhs.uk/live-well/exercise/running-and-aerobic-exercises/walking-for-health/

## Figures

[Figure 1:Box plot showing the steps assigned based on each diet type. 5](#_Toc124512514)

[Figure 2:Cumulative distribution function (CDF) of steps assigned by all markers. 6](#_Toc124512515)

## Tables

[Table 1: Summary of our independent variable showing all the categories. 7](#_Toc124512522)

[Table 2: Results after running the Shapiro Test for normality of the data. 7](#_Toc124512523)

[Table 3: P-values resulting from the pairwise t-test with the holm adjustment comparing the difference of means in the steps taken by an individual. 7](#_Toc124512524)