STAT-614 Final Report: Statistical Analysis of Diet Dataset

Report Prepared By:

Aditya Thakur

at4415@rit.edu

December 12, 2021

Table of Contents

Table of Figures

1.0 Introduction

This report summarizes all of the statistical analysis which is done on the diet dataset provided by University of Sheffield and publicly available on the Internet for teaching purposes. The diet dataset sample consists of 78 observations of individuals using one of three diets. The variables in the dataset include the primary key (named as Person), the individual's age, the individual's gender, the individual's height, the diet plan followed by the individual and the individual's weight before and after following the diet plan for 6 weeks. Out of these, the variables Gender and Diet have been used as factors. These are categorical variables. There is no response variable in the dataset. However, a new continuous response variable has been created which is the Weight loss computed by the difference between the individual's weight before and after following diet plan.

The aim of this statistical analysis is to answer the following questions.

i. Does following a diet plan really lead to weight loss?
ii. Is one of the diet plans more effective than the others?
iii. What is the effect of the individual's gender and diet plan followed on the individual's weight loss?

These questions are considered by dietitians while recommending a particular effective diet plan to the individuals who follow sedentary lifestyles. They recommend the diet based on the physical characteristics of the individual and then observe over a course of time if the diet plan is helping the concerned person with losing weight.

Question 1.1.i. has been answered by using a right tailed paired t-test while the questions 1.1.ii. and 1.1.iii. have been answered using a two factor ANOVA test.

2.0 Statistical Analysis

2.0.1 Numerical Summary of Response variable

The below Fig 2.0.1.1 shows the distribution of the response variable and the summary statistics for the weight loss are shown in the Table 2.0.1.1

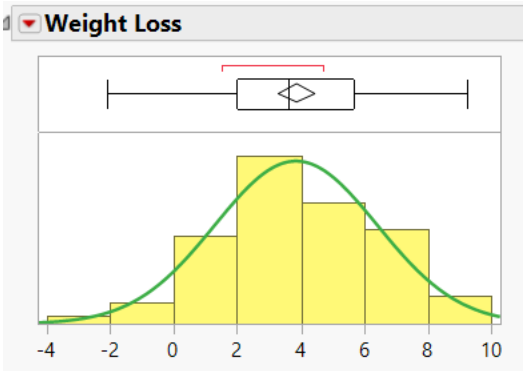

Fig 2.0.1.1. Histogram of Weight Loss

Summary Statistics	
Mean	3.8448718
Std Dev	2.5514777
Std Err Mean	0.288898
Upper 95% Mean	4.4201411
Lower 95% Mean	3.2696025
N	78

Table 2.0.1.1. Summary Statistics

2.0.2 Right-tailed paired t-test for Weight Loss column.

Fig 2.0.2.1 Paired t-test for response variable

2.0.2.1. Claim

The claim is that following a diet plan leads to weight loss.

2.0.2.2 Parameter of interest

The parameter of interest here is the mean weight loss (μ_D)

2.0.2.3 Hypotheses

Null Hypothesis $H_0 : \mu_D = 0$ (No weight loss after following diet plan)

Alternative Hypothesis $H_1 : \mu_D > 0$ (There is a weight loss after following diet plan)

2.0.2.4 Level of Significance

Assumed a 5% level of significance for this scenario ($\alpha = 0.05$). Used JMP Pro Matched Pairs functionality to perform this analysis. The results are shown in Fig 2.0.2.1

2.0.2.5 Test Statistic

From Fig 2.0.2.1 , the test statistic is the t-Ratio provided by JMP.

So, test statistic = 13.3088

2.0.2.6 p-value

From Fig 2.0.2.1 , since this is a right tailed test, we will select Prob > t

p-value < 0.0001

2.0.2.7 Decision

Since p-value ≤ α , our decision will be to Reject H_0

2.0.2.8 Validating Decision using confidence interval

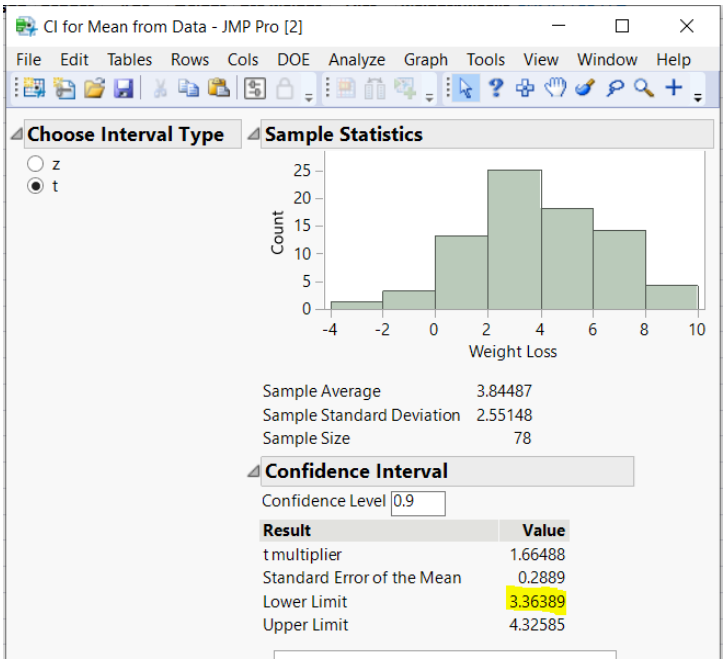

Fig 2.0.2.2 Lower confidence bound for mean weight loss

Since this is a right tailed test with 5% level of significance, we will create a 90% lower confidence bound for μ_D . This result is shown in Fig 2.0.2.2. As we can see, the lower bound is 3.3639 which is above our hypothesized mean 0. Hence, the decision to reject H_0 is valid.

2.0.3 Two Factor ANOVA for Diet and Gender columns

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
gender	1	1	1.285141	0.2383	0.6269
Diet	2	2	55.686508	5.1638	0.0080*
gender*Diet	2	2	40.919962	3.7945	0.0271*

Table 2.0.3.1 Two Factor ANOVA using Gender and Diet columns

2.0.3.1 Claim

The claim is that one of the diet plans is more effective than the others

2.0.3.2 Hypotheses

Null Hypothesis H_0: $\mu_{Diet1} = \mu_{Diet2} = \mu_{Diet3}$

Alternative Hypothesis H_a: At least one of the means differ

2.0.3.3 Level of Significance

Assumed a 5% level of significance for this scenario ($\alpha = 0.05$). Used JMP Pro Fit Model functionality to perform this analysis. The response variable (Y) is the weight loss and the factors are the Diet, Gender and their interactions. The results are shown in Table 2.0.3.1

2.0.3.4 Test Statistic

From Table 2.0.3.1, the test statistic for the Diet is F-Ratio 5.1638

2.0.3.5 p-value

From Table 2.0.3.1, the p-value for the Diet factor is 0.008

2.0.3.6 Decision

Since p-value $\leq \alpha$, our decision will be to Reject H_0

2.0.3.7 Additional Observations

From Table 2.0.3.1, we can see that the p-value for the Gender column is large and the p-value for the interactions between Gender and Diet is small.

2.0.3.8 Factor Profiling and Estimations

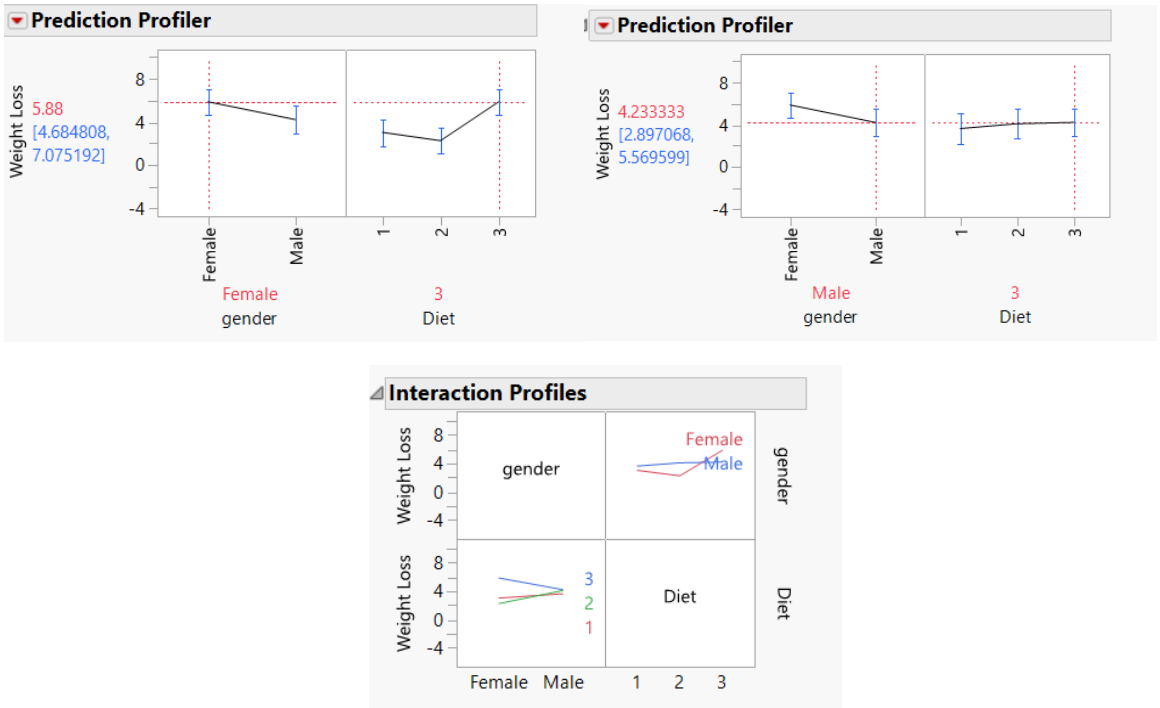

Fig 2.0.3.1 Prediction Profiler and Interaction Plot for Two Way ANOVA

From the above Fig 2.0.3.1, we can see the profiler shows that following diet plan 3 leads to a weight loss of 5.88 kg in case of females and 4.23 kg in case of males and the interaction plot shows that there is a high weight loss after following diet 3.

2.0.3.9 Validating Assumptions

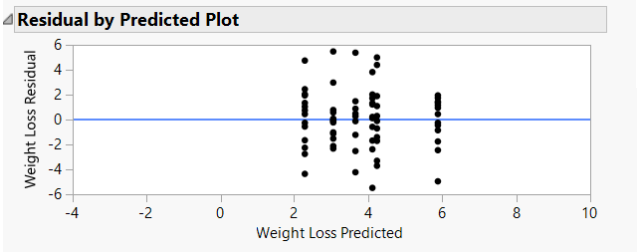

Fig 2.0.3.2 Residuals v/s Predicted Values plot

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	0.1867	2	75	0.8301
Brown-Forsythe	0.6257	2	75	0.5377
Levene	0.6589	2	75	0.5204
Bartlett	0.1687	2	.	0.8447

Table 2.0.3.2 Test for unequal variances

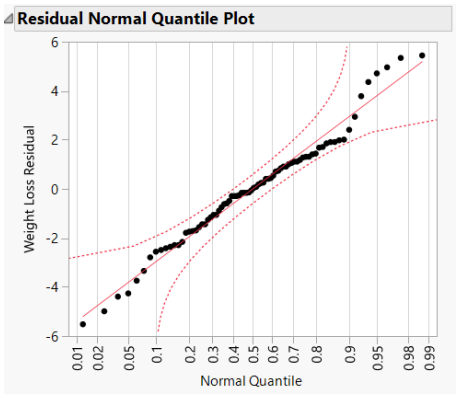

Fig 2.0.3.3 Normal Quantile Plot of residuals

From Fig 2.0.3.2, it seems that the equal variances assumption is valid. Also, from the Table 2.0.3.2, the test for unequal variances is not significant. Hence, we can assume equal variances. Also, from Fig 2.0.3.3 since the quantile plot of residuals is a straight line, we can assume that the residuals are normally distributed. Hence, normal assumption is also valid.

3.0 Conclusions

(i) From the section 2.0.2.7, since we rejected the null hypothesis of the paired t-test , we can conclude at 5% level of significance, there is sufficient evidence that following a diet plan does lead to weight loss. This answers the question 1.1.i.

(ii) From the section 2.0.3.6, since we rejected the null hypothesis of the two factor ANOVA , we can conclude at 5% level of significance, there is sufficient evidence that one of the diet plan is more effective than the others. Similarly, from section 2.0.3.7, we can conclude that gender does not have an impact on weight loss but the interactions between Gender and Diet do. This answers the questions 1.1.ii. and 1.1.iii.

4.0 Recommendations

(i) For individuals following sedentary lifestyles, it is recommended to follow a diet plan as there is sufficient evidence that it helps in weight loss.

(ii) Irrespective of the gender, for the best results in terms of weight loss, it is recommended for individuals to follow diet plan number 3.

5.0 References

https://www.sheffield.ac.uk/polopoly_fs/1.937194!/file/Diet_data_description.doc