

# STAT-614 PROJECT PROPOSAL

Aditya Thakur  
November 29, 2021  
at4415@rit.edu

## PROJECT TITLE

Statistical Analysis of Diet Dataset using JMP Pro

## DATASET DESCRIPTION

The proposed project is to perform a statistical analysis of the Diet dataset provided by University of Sheffield for teaching purposes. The dataset is publicly available online at the link [https://www.sheffield.ac.uk/polopoly\\_fs/1.937195!/file/Diet\\_R.csv](https://www.sheffield.ac.uk/polopoly_fs/1.937195!/file/Diet_R.csv). 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 will be used as factors. These are categorical variables. There is no response variable in the dataset. However, I will be creating a new continuous response variable which will be the weight loss and it will be computed by the difference between the individual's weight before and after following diet plan.

## OBJECTIVES

The questions that I will be looking to answer through the statistical analysis are as follows:

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

I am trying to answer these questions as these insights are the ones which are often considered by dietitians while recommending a particular effective diet plan to 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 indeed helping the concerned person with losing weight.

## SOLUTION

First of all, the dataset contains two null values in the Gender column. So I will impute the null values with the mode of the Gender column (since it is categorical).

- To answer the first question, I will use a two sample paired t-test in which I will consider the created weight loss column as the column of mean differences
- To answer the second question, I will use a one factor ANOVA test with the diet plan as the involved factor.
- To answer the third question, I will use a two factor ANOVA test with the individual's gender and diet plan as factors as well the interactions among them.