DATA 603 L01 05 Final Report

Group L01-05 Members:

1. Steen Rasmussen - 30097313
2. Aaron Gelfand - 10032214
3. Jackson Meier - 30095291
4. Harpreet Saini - 30271048
5. Venkateshwaran Balu Soundararajan - 30239509

1 Introduction

For our project, we will be analyzing various factors that influence weight change using a multilinear regression model. Our dataset contains information relating to diet, physical activity and various lifestyle habits that are known to influence weight fluctuations. Through our analysis, we hope to identify what factors should be limited to prevent weight gain, and what factors should be encouraged to promote weight loss. Through this analysis, we hope to provide insight and various suggestions as to how to properly manage weight from both individual and public health perspectives. Obesity has become an extremely important topic in recent years, causing more health problems than we have ever seen before (Twells et al. 2014). Our main objective will be to predict weight changes based on these lifestyle and dietary variables. Through the use of our multilinear regression model, we will identify the significant variables that influence weight change, eliminate insignificant variables, resulting in a consistent and reliable framework to determine weight change. Other studies in this area identify lifestyle habits such as caloric intake, physical activity and quantity of sleep as important factors (National Institute of Diabetes and Digestive and Kidney Diseases, n.d.).

2 Methodology

2.1 Data

The dataset [2] is publicly available as open-source data on Kaggle, it was provided by an external source and was not collected or created by any members of our group. The team is using this dataset only for analysis and research purposes, with no changes in its original data collection processes. The dataset is free to access publicly without any licensing and permission required, but the credentials are mentioned.

This dataset comprises information from 100 participants, focusing on demographics, dietary habits, physical activity levels, and lifestyle factors to predict weight change over time. Key features include age, gender, current weight, daily caloric intake, macronutrient breakdown, sleep quality, and stress levels. Based on this we aimed to analyze how these variables interact and influence weight fluctuations, providing a valuable resource for researchers and practitioners in nutrition and health.

Based on the analysis we have categorized the qualitative and quantitative variables as below based on the data set

S. No	Variable Name	Variable Type	Comments	Possible Values
1	Participant ID	Quantitative	Auto Increment	
2	Weight Change (lbs)	Quantitative	Response	
3	Age	Quantitative	Predictor	
4	Current Weight (lbs)	Quantitative	Predictor	
5	BMR (Calories)	Quantitative	Predictor	
6	Daily Calories Consumed	Quantitative	Predictor	
7	Daily Caloric Surplus/Deficit	Quantitative	Predictor	
8	Duration (weeks)	Quantitative	Predictor	
9	Final Weight (lbs)	Quantitative	Predictor	
10	Gender	Categorical	Predictor	2 Level: M, F
11	Physical Activity Level	Categorical	Predictor	5 Level: Physical Activity Level, Sedentary, Very Active, Lightly Active, Physical Activity Level
12	Sleep Quality	Categorical	Predictor	4 Level: Excellent, Good, Fair, Poor
13	Stress Level	Categorical	Predictor	9 Level: Range (1,9)

2.2 Approach

The primary objective of this project is to identify the key factors that significantly influence weight changes in the human body. To achieve this, we will build multiple models based on the following guiding questions:

Guiding Question 1: "Analyze how do gender, Physical Activity Level, Sleep Quality, and Stress Level, Age, Current Weight, BMR, Daily Calories Consumed, Daily Caloric Surplus/Deficit and Duration of the 100 participants collectively influence weight change, and are there significant interactions between these factors that modify their effects on weight change?"

Guiding Question 2: "Analyze how do age, basal metabolic rate, daily caloric intake, and caloric surplus or deficit affect weight change over the program's duration? Are there any combinations of these factors that are more strongly connected with weight change?"

Guiding Question 3: " Analyze how gender, physical activity level, sleep quality, and stress level affect weight change in adults, both individually and in combination?" For example, does increased physical activity lower weight more effectively in low-stress situations, or does excellent sleep quality play a larger impact at specific stress and activity levels?"

Hence based on the interpretation of these model we will determine the highest factors that play a vital role to reflect the weight changes in the human body and derive the best fit model for weight prediction

2.3 Workflow

Workflow Task List:

Step 1: Data Loading and Wrangling The first step in any data analysis project is to load the dataset and perform necessary data wrangling. This involves cleaning, preprocessing, and transforming the data to ensure accuracy and reliability. Using R, we can load the dataset into a data frame and apply various transformations to prepare the data for analysis.

Step 2: Variable Identification and Removal Before performing any regression tests, it's crucial to identify and remove variables that do not support the modelling process. This includes auto-increment variables, index variables, and any metadata information. Removing these variables helps in focusing on the relevant predictors and improves the model's performance.

Step 3: Building the First Order Model With the cleaned dataset, we build a first-order model using linear regression. This model helps in understanding the relationship between the dependent variable and the independent variables. We then perform individual T-tests to evaluate the significance of each predictor.

Step 4: Residual Analysis and Multicollinearity Check Residual analysis is essential to check the assumptions of regression. We plot the residuals to ensure they are randomly distributed. Additionally, we use the Variance Inflation Factor (VIF) method to test for multicollinearity among the predictors. Variables with high VIF values are eliminated to improve the model.

Step 5: Building the Adjusted Model After removing variables with high multicollinearity, we build an adjusted model. This refined model is subjected to individual T-tests to evaluate the significance of the remaining predictors. This step ensures that only the most relevant variables are included in the model.

Step 6: Model Selection Procedures Before proceeding to the interaction model, we run VIF again with the adjusted model to verify multicollinearity. We also apply model selection procedures, such as the All-Possible-Regressions Selection Procedures -Adjusted R^2 or RSE Criterion, Mallows Cp Criterion and Akaike information criterion (AIC) to identify the most significant predictors from the first-order model.

Step 7: Building the Interaction Model Next, we build an interaction model to explore the interactions between predictors. We perform hypothesis tests to identify significant interaction terms and adjust the model accordingly. This step helps in capturing the combined effect of multiple predictors on the dependent variable.

Step 8: Testing Linearity of the Interaction Model To ensure the linearity of the final interaction model, we interpret the residual plot. This involves checking for patterns in the residuals that might indicate non-linearity. If necessary, we adjust improve the model's linearity.

Step 9: Improving Linearity with Transformations To further improve the linearity of the model, we add polynomial terms or apply log transformations to potential predictors. These transformations help in capturing non-linear relationships and enhance the model's efficiency.

Step 10: Testing for Homoscedasticity We run the Breusch-Pagan test to check for homoscedasticity, which ensures that the variance of the residuals is constant across all levels of the independent variables. This step is crucial for validating the assumptions of linear regression.

Step 11: Verifying Normality Using the Shapiro-Wilks test, we verify the normality of the residuals. This test helps in confirming that the residuals are normally distributed, which is an important assumption for linear regression models.

Step 12: Finalizing the Model, Making Interpretations and Predictions After verifying all assumptions and making necessary adjustments, we finalize the model and make interpretations of the model. The final step involves using the model to make predictions on new data. This step demonstrates the practical application of the model and its ability to provide actionable insights.

2.4 Contributions

Jackson: Developed Project Introduction and Defined objectives, Interpretation and Conclusion

Venkateshwaran: Expanded the Project Methodology, Categorized the dataset variables and defined the workflow/Task list for the models being developed and Tested

Steen: Model with Quantitative variables - Responsible for building the full model with interaction terms that will focus on exploring only the quantitative variables (Age, Current Weight, BMR, Daily Calories Consumed, Daily Caloric Surplus/Deficit, Weight Change, Duration, and Final Weight).

Harpreet: Model with Qualitative Variables – Concentrate on assessing the roles played by these qualitative variables (Gender, Physical Activity Level, Sleep Quality, and Stress Level) in the model, detailing how they modify or account for differences in the dependent variable.

Aaron: Model with all variables - Responsible for building the full model with interaction terms that will focus on exploring relationships among both qualitative and quantitative variables. A multiple regression model with interaction terms will allow them to assess not only the individual impact of each predictor on the outcome variable but also to examine how the effect of one predictor might change depending on the levels of another (e.g., stress level affecting the relationship between physical activity level and weight change).

3 Main Results of the Analysis

To demonstrate the importance of including all variable types in our predictive model, we decided to create and compare three models:

1. A quantitative model
2. A qualitative model
3. A model containing quantitative and qualitative variables

Before creating any of our models, we made sure the appropriate libraries were uploaded, and the dataset was uploaded.

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2

## Loading required package: carData

## Warning: package 'lmtest' was built under R version 4.4.2

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

##
## Attaching package: 'olsrr'

## The following object is masked from 'package:MASS':
##
##     cement

## The following object is masked from 'package:datasets':
##
##     rivers

## Warning: package 'Ecdat' was built under R version 4.4.2

## Loading required package: Ecfun
```

```
## Warning: package 'Ecfun' was built under R version 4.4.2

##
## Attaching package: 'Ecfun'

## The following object is masked from 'package:base':
##
##     sign

##
## Attaching package: 'Ecdat'

## The following object is masked from 'package:MASS':
##
##     SP500

## The following object is masked from 'package:carData':
##
##     Mroz

## The following object is masked from 'package:datasets':
##
##     Orange

##   Participant.ID Age Gender Current.Weight..lbs. BMR..Calories.
## 1              1  56      M                228.4         3102.3
## 2              2  46      F                165.4         2275.5
## 3              3  32      F                142.8         2119.4
## 4              4  25      F                145.5         2181.3
## 5              5  38      M                155.5         2463.8
## 6              6  56      F                152.9         2100.6
##   Daily.Calories.Consumed Daily.Caloric.Surplus.Deficit Weight.Change..lbs.
## 1                  3916.0                         813.7              0.2000
## 2                  3823.0                        1547.5              2.4000
## 3                  2785.4                         666.0              1.4000
## 4                  2587.3                         406.0              0.8000
## 5                  3312.8                         849.0              2.0000
## 6                  2262.4                         161.9            -12.5135
##   Duration..weeks. Physical.Activity.Level Sleep.Quality Stress.Level
## 1                1               Sedentary     Excellent            6
## 2                6             Very Active     Excellent            6
## 3                7               Sedentary          Good            3
## 4                8               Sedentary          Fair            2
## 5               10          Lightly Active          Good            1
## 6                9               Sedentary          Poor            6
##   Final.Weight..lbs.
## 1              228.6
## 2              167.8
## 3              144.2
## 4              146.3
## 5              157.5
## 6              140.4
```

If there is an issue uploading the dataset from the github link that has been used, a copy of the dataset has been included in our submission, that can be manually uploaded.

3.1 Quantitative Model

To construct our quantitative model, we begin by inspecting our dataset to determine which variables should be included.

```
##  [1] "Participant.ID"                 "Age"
##  [3] "Gender"                         "Current.Weight..lbs."
##  [5] "BMR..Calories."                 "Daily.Calories.Consumed"
##  [7] "Daily.Caloric.Surplus.Deficit" "Weight.Change..lbs."
##  [9] "Duration..weeks."               "Physical.Activity.Level"
## [11] "Sleep.Quality"                  "Stress.Level"
## [13] "Final.Weight..lbs."
```

Variables under consideration:

[1] "Participant.ID"
[2] "Age"
[3] "Gender"
[4] "Current.Weight..lbs."
[5] "BMR..Calories."
[6] "Daily.Calories.Consumed"
[7] "Daily.Caloric.Surplus.Deficit" [8] "Weight.Change..lbs."
[9] "Duration..weeks."
[10] "Physical.Activity.Level"
[11] "Sleep.Quality"
[12] "Stress.Level"
[13] "Final.Weight..lbs."

We will begin by excluding Participant.ID and any qualitative variables, leaving us with:

[1] "Age"
[2] "Current.Weight..lbs."
[3] "BMR..Calories."
[4] "Daily.Calories.Consumed"
[5] "Daily.Caloric.Surplus.Deficit" [6] "Weight.Change..lbs."
[7] "Duration..weeks."
[8] "Final.Weight..lbs."

We will then test that at least one of our predictors is significant using the following hypothesis:

$H_0 : \beta_1 = \beta_2 = ... = \beta_i = 0$ (i=1,2,...,p)
$H_a :$ At least one $\beta_i \neq 0$ (i=1,2,...,p)

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ Age + Current.Weight..lbs. +
##     BMR..Calories. + Daily.Calories.Consumed + Daily.Caloric.Surplus.Deficit +
##     Duration..weeks. + Final.Weight..lbs., data = weight)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.054593 -0.006195 -0.000756  0.004415  0.047691
##
## Coefficients:
##                                Estimate Std. Error   t value Pr(>|t|)
## (Intercept)                   1.454e-02  2.121e-02     0.685    0.495
```

```
## Age                            -2.411e-05  2.024e-04    -0.119    0.905
## Current.Weight..lbs.           -9.996e-01  3.647e-04 -2740.998   <2e-16 ***
## BMR..Calories.                 -3.607e-02  5.307e-02    -0.680    0.498
## Daily.Calories.Consumed         3.607e-02  5.307e-02     0.680    0.498
## Daily.Caloric.Surplus.Deficit -3.606e-02  5.307e-02    -0.680    0.499
## Duration..weeks.               -3.329e-05  6.211e-04    -0.054    0.957
## Final.Weight..lbs.              9.996e-01  2.915e-04  3429.410   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02119 on 92 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 1.746e+06 on 7 and 92 DF,  p-value: < 2.2e-16
```

The predictors Current.Weight..lbs. and Final.Weight..lbs. are highly significant but when used in relation to Weight.Change..lbs. are too correlated. Hence, we will remove them because weight change is calculated using the current weight and the final weight.

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ Age + BMR..Calories. + Daily.Calories.Consumed +
##     Daily.Caloric.Surplus.Deficit + Duration..weeks., data = weight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -30.924  -2.704   1.683   4.674   9.202
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     3.24270    6.78058   0.478    0.634
## Age                             0.02333    0.06373   0.366    0.715
## BMR..Calories.                -13.79019   18.71243  -0.737    0.463
## Daily.Calories.Consumed        13.78798   18.71224   0.737    0.463
## Daily.Caloric.Surplus.Deficit -13.78740   18.71234  -0.737    0.463
## Duration..weeks.               -0.26693    0.21795  -1.225    0.224
##
## Residual standard error: 7.495 on 94 degrees of freedom
## Multiple R-squared:  0.03738,    Adjusted R-squared:  -0.01382
## F-statistic: 0.7301 on 5 and 94 DF,  p-value: 0.6026
```

After removing Current.Weight..lbs. and Final.Weight..lbs, we see no statistically significant predictors and we have an adjusted R^2 of -0.01382 which suggests that the remaining predictors have limited explanatory value. This low adjusted R^2 indicates that more refinement is needed.

As we are still uncertain whether the model has multicollinearity, we will now test the remaining variables for that.

```
##                           Age                BMR..Calories.
##                  1.068859e+00                  8.195700e+07
##       Daily.Calories.Consumed Daily.Caloric.Surplus.Deficit
##                  1.625962e+08                  8.519425e+07
##              Duration..weeks.
##                  1.034519e+00
```

```
##
## Call:
## imcdiag(mod = quant_weight_model_take2, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                                          VIF detection
## Age                              1.068900e+00         0
## BMR..Calories.                   8.195700e+07         1
## Daily.Calories.Consumed          1.625962e+08         1
## Daily.Caloric.Surplus.Deficit 8.519425e+07         1
## Duration..weeks.                 1.034500e+00         0
##
## Multicollinearity may be due to BMR..Calories. Daily.Calories.Consumed Daily.Caloric.Surplus.Deficit
##
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test
##
## ===================================
```

After checking the VIF, we have the result: Age = 1.068859, BMR..Calories. = 8.185700×10^7, Daily.Calories.Consumed = 1.625962×10^8, Daily.Caloric.Surplus.Deficit = 8.519425×10^7, and Duration..weeks. = 1.034519. Values greater than 10 suggest severe multicollinearity, hence BMR..Calories., Daily.Calories.Consumed, and Daily.Caloric.Surplus.Deficit are caught in VIF detection and need to be addressed. Based on our data dictionary we can see that Daily.Caloric.Surplus.Deficit is the difference between Daily.Calories.Consumed and BMR..Calories., therefore we have chosen to keep Daily.Caloric.Surplus.Deficit while removing the other two variables. Leaving us with the model below:

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ Age + Daily.Caloric.Surplus.Deficit +
##     Duration..weeks., data = weight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -31.819  -2.310   1.403   4.602   9.156
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -2.6238347  3.4349578  -0.764    0.447
## Age                            0.0298644  0.0618788   0.483    0.630
## Daily.Caloric.Surplus.Deficit  0.0006742  0.0020356   0.331    0.741
## Duration..weeks.              -0.2835862  0.2155103  -1.316    0.191
##
## Residual standard error: 7.475 on 96 degrees of freedom
## Multiple R-squared:  0.02201,    Adjusted R-squared:  -0.008555
## F-statistic: 0.7201 on 3 and 96 DF,  p-value: 0.5424
```

From our output we can see that the p-value of our model is quite large, at 0.5424, and none of our predictors our significant. We can attempt a stepwise selection to see if any predictors should be included.

NOTE The codeline below is only included to demonstrate the code that was used. The output returns an error indicating that none of the variables are appropriate to select.

```r
#stepmod=ols_step_both_p(quant_weight_model_take3, p_enter=0.05,p_remove=0.3,details=TRUE)
```

Based on our output, the large p-value of our model, and the fact that no predictors have a p-value < 0.05, we conclude that this model is not significantly different than a model with no predictors. This suggests that we should attempt other modelling approaches, such as a generalized additive model. However, as this is not covered in the scope of our course, we will simply conclude that using only the quantitative variables is not a good method for determining weight change given our dataset.

3.2 Qualitative Model

```
##   Participant.ID Age Gender Current.Weight..lbs. BMR..Calories.
## 1              1  56      M                228.4         3102.3
## 2              2  46      F                165.4         2275.5
## 3              3  32      F                142.8         2119.4
## 4              4  25      F                145.5         2181.3
## 5              5  38      M                155.5         2463.8
## 6              6  56      F                152.9         2100.6
##   Daily.Calories.Consumed Daily.Caloric.Surplus.Deficit Weight.Change..lbs.
## 1                  3916.0                         813.7              0.2000
## 2                  3823.0                        1547.5              2.4000
## 3                  2785.4                         666.0              1.4000
## 4                  2587.3                         406.0              0.8000
## 5                  3312.8                         849.0              2.0000
## 6                  2262.4                         161.9            -12.5135
##   Duration..weeks. Physical.Activity.Level Sleep.Quality Stress.Level
## 1                1               Sedentary     Excellent            6
## 2                6             Very Active     Excellent            6
## 3                7               Sedentary          Good            3
## 4                8               Sedentary          Fair            2
## 5               10          Lightly Active          Good            1
## 6                9               Sedentary          Poor            6
##   Final.Weight..lbs.
## 1              228.6
## 2              167.8
## 3              144.2
## 4              146.3
## 5              157.5
## 6              140.4
```

Meet our, 100 participants who has decided to embark on a transformative journey to improve their health and well-being. They are determined to contribute to body test which is going to be really helpful in the health department to understand how various lifestyle factors influence the body weight.

For the purposes of this section, we will be creating a model that focuses on qualitative variables, starting with the following variables -

1. Gender- M or F
2. Physical.Activity.Level
3. Sleep.Quality
4. Stress.Level

Both genders commits to track their progress with how their physical activity level is throughout that period of time. Additionally, mental aspects were recorded such as their sleep quality and stress levels.

We can determine whether our model contains any significant predictors using the hypothesis:

$H_0 : \beta_1 = \beta_2 = ... = \beta_i = 0$ (i=1,2,...,p)
$H_a :$ At least one $\beta_i \neq 0$ (i=1,2,...,p)

#Full Model

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ factor(Gender) + factor(Physical.Activity.Level) +
##     factor(Sleep.Quality) + factor(Stress.Level), data = weight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.313  -1.369  -0.074   2.093  12.709
##
## Coefficients:
##                                                Estimate Std. Error t value
## (Intercept)                                      1.7189     1.8081   0.951
## factor(Gender)M                                 -0.8365     0.9712  -0.861
## factor(Physical.Activity.Level)Moderately Active  0.0540     1.2103   0.045
## factor(Physical.Activity.Level)Sedentary        -0.6371     1.2509  -0.509
## factor(Physical.Activity.Level)Very Active       1.4788     1.2152   1.217
## factor(Sleep.Quality)Fair                        0.5978     1.3964   0.428
## factor(Sleep.Quality)Good                        1.1141     1.4241   0.782
## factor(Sleep.Quality)Poor                       -8.1191     1.2927  -6.281
## factor(Stress.Level)2                            0.2977     1.7632   0.169
## factor(Stress.Level)3                            1.2671     1.7525   0.723
## factor(Stress.Level)4                            1.5876     2.0526   0.773
## factor(Stress.Level)5                            1.8938     1.8458   1.026
## factor(Stress.Level)6                           -0.2817     1.7490  -0.161
## factor(Stress.Level)7                           -0.2628     1.8777  -0.140
## factor(Stress.Level)8                          -10.8574     1.8308  -5.930
## factor(Stress.Level)9                           -9.1820     1.8864  -4.867
##                                                Pr(>|t|)
## (Intercept)                                       0.344
## factor(Gender)M                                   0.392
## factor(Physical.Activity.Level)Moderately Active   0.965
## factor(Physical.Activity.Level)Sedentary          0.612
## factor(Physical.Activity.Level)Very Active        0.227
## factor(Sleep.Quality)Fair                         0.670
## factor(Sleep.Quality)Good                         0.436
## factor(Sleep.Quality)Poor                      1.42e-08 ***
## factor(Stress.Level)2                             0.866
## factor(Stress.Level)3                             0.472
## factor(Stress.Level)4                             0.441
## factor(Stress.Level)5                             0.308
## factor(Stress.Level)6                             0.872
## factor(Stress.Level)7                             0.889
## factor(Stress.Level)8                          6.50e-08 ***
## factor(Stress.Level)9                          5.23e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.159 on 84 degrees of freedom
```

```
## Multiple R-squared:  0.7351, Adjusted R-squared:  0.6878
## F-statistic: 15.54 on 15 and 84 DF,  p-value: < 2.2e-16
```

Based on our output, we can see that there is at least on predictor with a p-value < 0.05, allowing us to reject the null hypothesis and conclude that at least one predictor is significant.

To determine which predictors are significant, we can test the hypothesis:

$H_0 : \beta_i = 0$ (i=1,2,...,p)
$H_a : \beta_i \neq 0$ (i=1,2,...,p)

From our output above we can see that the predictors Sleep.Quality and Stress.Level contain at least one factor with a p-value < 0.05. We therefore reject the null hypothesis and conclude that Sleep.Quality and Stress.Level significantly affect weight change and should be included in our model.

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ factor(Sleep.Quality) + factor(Stress.Level),
##     data = weight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.0668  -1.5978  -0.1951   2.0673  14.2170
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.66777    1.64232   1.015    0.313
## factor(Sleep.Quality)Fair   0.66673    1.37884   0.484    0.630
## factor(Sleep.Quality)Good   0.86678    1.41286   0.613    0.541
## factor(Sleep.Quality)Poor  -7.89997    1.28104  -6.167 2.06e-08 ***
## factor(Stress.Level)2      -0.16713    1.69144  -0.099    0.922
## factor(Stress.Level)3       0.72075    1.68828   0.427    0.670
## factor(Stress.Level)4       1.12429    2.01231   0.559    0.578
## factor(Stress.Level)5       1.45804    1.78473   0.817    0.416
## factor(Stress.Level)6      -0.05374    1.71822  -0.031    0.975
## factor(Stress.Level)7      -0.14142    1.82197  -0.078    0.938
## factor(Stress.Level)8     -11.34225    1.77596  -6.387 7.78e-09 ***
## factor(Stress.Level)9      -9.37917    1.86760  -5.022 2.65e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.145 on 88 degrees of freedom
## Multiple R-squared:  0.7244, Adjusted R-squared:   0.69
## F-statistic: 21.03 on 11 and 88 DF,  p-value: < 2.2e-16
```

From our output we can see that our refined model has a slightly better **Adjusted R-squared** (0.69) than the full model (0.6878), indicating an improvement in model simplicity and explanatory power.

The variables **factor(Sleep.Quality)Poor**, **factor(Stress.Level)8**, and **factor(Stress.Level)9** remain highly significant and have the greatest impact on weight change.

Ton ensure no issues with multicollinearity we can run a VIF test on our model

```
##                           GVIF Df GVIF^(1/(2*Df))
## factor(Sleep.Quality) 1.227473  3        1.034750
## factor(Stress.Level)  1.227473  8        1.012892
```

```
##
## Call:
## imcdiag(mod = weight_refined_model, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                               VIF detection
## factor(Sleep.Quality)Fair 2.0186         0
## factor(Sleep.Quality)Good 1.9939         0
## factor(Sleep.Quality)Poor 2.2506         0
## factor(Stress.Level)2     2.0051         0
## factor(Stress.Level)3     1.9976         0
## factor(Stress.Level)4     1.5345         0
## factor(Stress.Level)5     1.8152         0
## factor(Stress.Level)6     1.9436         0
## factor(Stress.Level)7     1.7391         0
## factor(Stress.Level)8     1.7974         0
## factor(Stress.Level)9     1.6628         0
##
## NOTE:  VIF Method Failed to detect multicollinearity
##
##
## 0 --> COLLINEARITY is not detected by the test
##
## ===================================
```

From our output we can see that the VIF is less than 5 so all predictors are kept as there is no indication of significant multicollinearity.

Now that we have concluded the absence of multicollinearity, we can begin strengthening our model. To begin, we will look at adding interaction terms. Interaction terms are useful when we suspect that the effect of one predictor on the response variable depends on the level of another predictor.

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ (factor(Sleep.Quality) + factor(Stress.Level))^2,
##     data = weight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.509  -1.190   0.000   1.251  16.775
##
## Coefficients: (2 not defined because of singularities)
##                                                     Estimate Std. Error t value
## (Intercept)                                          0.90000    4.36708   0.206
## factor(Sleep.Quality)Fair                            2.00000    5.04267   0.397
## factor(Sleep.Quality)Good                            0.96667    5.04267   0.192
## factor(Sleep.Quality)Poor                           -6.86346    4.88254  -1.406
## factor(Stress.Level)2                                1.20000    5.04267   0.238
## factor(Stress.Level)3                               -0.30000    5.04267  -0.059
## factor(Stress.Level)4                                4.15260    3.78200   1.098
## factor(Stress.Level)5                                0.85000    5.34856   0.159
## factor(Stress.Level)6                                1.50000    4.88254   0.307
```

```
## factor(Stress.Level)7                              -0.40000    5.34856  -0.075
## factor(Stress.Level)8                              -4.57874    6.17598  -0.741
## factor(Stress.Level)9                             -12.20581    3.08799  -3.953
## factor(Sleep.Quality)Fair:factor(Stress.Level)2   -3.14000    5.96657  -0.526
## factor(Sleep.Quality)Good:factor(Stress.Level)2   -0.60667    5.96657  -0.102
## factor(Sleep.Quality)Poor:factor(Stress.Level)2    2.14010    7.01910   0.305
## factor(Sleep.Quality)Fair:factor(Stress.Level)3    1.80000    7.13141   0.252
## factor(Sleep.Quality)Good:factor(Stress.Level)3    0.36667    6.17598   0.059
## factor(Sleep.Quality)Poor:factor(Stress.Level)3    1.89297    5.73767   0.330
## factor(Sleep.Quality)Fair:factor(Stress.Level)4   -4.25260    5.49512  -0.774
## factor(Sleep.Quality)Good:factor(Stress.Level)4   -4.11927    5.19786  -0.792
## factor(Sleep.Quality)Poor:factor(Stress.Level)4         NA         NA      NA
## factor(Sleep.Quality)Fair:factor(Stress.Level)5   -2.15000    6.67082  -0.322
## factor(Sleep.Quality)Good:factor(Stress.Level)5    1.68333    7.35089   0.229
## factor(Sleep.Quality)Poor:factor(Stress.Level)5    1.46086    6.04595   0.242
## factor(Sleep.Quality)Fair:factor(Stress.Level)6   -2.46667    6.04595  -0.408
## factor(Sleep.Quality)Good:factor(Stress.Level)6   -1.56667    6.30334  -0.249
## factor(Sleep.Quality)Poor:factor(Stress.Level)6   -2.00745    5.77710  -0.347
## factor(Sleep.Quality)Fair:factor(Stress.Level)7    0.06667    6.42817   0.010
## factor(Sleep.Quality)Good:factor(Stress.Level)7    3.33333    7.35089   0.453
## factor(Sleep.Quality)Poor:factor(Stress.Level)7   -0.37889    6.17598  -0.061
## factor(Sleep.Quality)Fair:factor(Stress.Level)8   -6.58104    7.35089  -0.895
## factor(Sleep.Quality)Good:factor(Stress.Level)8   -2.98419    7.35089  -0.406
## factor(Sleep.Quality)Poor:factor(Stress.Level)8   -9.31800    6.78890  -1.373
## factor(Sleep.Quality)Fair:factor(Stress.Level)9    4.63575    4.71699   0.983
## factor(Sleep.Quality)Good:factor(Stress.Level)9    5.04842    5.04267   1.001
## factor(Sleep.Quality)Poor:factor(Stress.Level)9         NA         NA      NA
##                                                   Pr(>|t|)
## (Intercept)                                       0.837357
## factor(Sleep.Quality)Fair                         0.692931
## factor(Sleep.Quality)Good                         0.848568
## factor(Sleep.Quality)Poor                         0.164499
## factor(Stress.Level)2                             0.812642
## factor(Stress.Level)3                             0.952740
## factor(Stress.Level)4                             0.276198
## factor(Stress.Level)5                             0.874216
## factor(Stress.Level)6                             0.759646
## factor(Stress.Level)7                             0.940611
## factor(Stress.Level)8                             0.461094
## factor(Stress.Level)9                             0.000191 ***
## factor(Sleep.Quality)Fair:factor(Stress.Level)2 0.600468
## factor(Sleep.Quality)Good:factor(Stress.Level)2 0.919321
## factor(Sleep.Quality)Poor:factor(Stress.Level)2 0.761404
## factor(Sleep.Quality)Fair:factor(Stress.Level)3 0.801513
## factor(Sleep.Quality)Good:factor(Stress.Level)3 0.952837
## factor(Sleep.Quality)Poor:factor(Stress.Level)3 0.742506
## factor(Sleep.Quality)Fair:factor(Stress.Level)4 0.441761
## factor(Sleep.Quality)Good:factor(Stress.Level)4 0.430913
## factor(Sleep.Quality)Poor:factor(Stress.Level)4       NA
## factor(Sleep.Quality)Fair:factor(Stress.Level)5 0.748245
## factor(Sleep.Quality)Good:factor(Stress.Level)5 0.819579
## factor(Sleep.Quality)Poor:factor(Stress.Level)5 0.809819
## factor(Sleep.Quality)Fair:factor(Stress.Level)6 0.684604
## factor(Sleep.Quality)Good:factor(Stress.Level)6 0.804484
```

```
## factor(Sleep.Quality)Poor:factor(Stress.Level)6 0.729334
## factor(Sleep.Quality)Fair:factor(Stress.Level)7 0.991757
## factor(Sleep.Quality)Good:factor(Stress.Level)7 0.651705
## factor(Sleep.Quality)Poor:factor(Stress.Level)7 0.951266
## factor(Sleep.Quality)Fair:factor(Stress.Level)8 0.373896
## factor(Sleep.Quality)Good:factor(Stress.Level)8 0.686083
## factor(Sleep.Quality)Poor:factor(Stress.Level)8 0.174544
## factor(Sleep.Quality)Fair:factor(Stress.Level)9 0.329306
## factor(Sleep.Quality)Good:factor(Stress.Level)9 0.320415
## factor(Sleep.Quality)Poor:factor(Stress.Level)9       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.367 on 66 degrees of freedom
## Multiple R-squared:  0.7705, Adjusted R-squared:  0.6558
## F-statistic: 6.716 on 33 and 66 DF,  p-value: 3.128e-11
```

The interaction model does not improve predictive performance or interpretability over the refined model. The refined model remains the best choice for explaining the relationship between weight change, Sleep Quality, and Stress Level.

Furthermore, the Adjusted R sqaure in our interaction model has decreased from our refined model (0.6558, down from 0.69), and the RSE has increased (4.367, up from 4.145). We also see that only one predictor has a p-value < 0.05. Based on all of this, we conclude that the refined model is better than the presented interaction model.

To further improve our model we can attempt the addition of higher order terms. **Note** Use of I(factor()^2) is for continuous variables where you want to square the variable, but it's not applicable to categorical variables. As here we have all categorical variables we will just do it like -

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ (factor(Sleep.Quality) + factor(Stress.Level))^3,
##     data = weight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.509  -1.190   0.000   1.251  16.775
##
## Coefficients: (2 not defined because of singularities)
##                                                  Estimate Std. Error t value
## (Intercept)                                       0.90000    4.36708   0.206
## factor(Sleep.Quality)Fair                         2.00000    5.04267   0.397
## factor(Sleep.Quality)Good                         0.96667    5.04267   0.192
## factor(Sleep.Quality)Poor                        -6.86346    4.88254  -1.406
## factor(Stress.Level)2                             1.20000    5.04267   0.238
## factor(Stress.Level)3                            -0.30000    5.04267  -0.059
## factor(Stress.Level)4                             4.15260    3.78200   1.098
## factor(Stress.Level)5                             0.85000    5.34856   0.159
## factor(Stress.Level)6                             1.50000    4.88254   0.307
## factor(Stress.Level)7                            -0.40000    5.34856  -0.075
## factor(Stress.Level)8                            -4.57874    6.17598  -0.741
## factor(Stress.Level)9                           -12.20581    3.08799  -3.953
## factor(Sleep.Quality)Fair:factor(Stress.Level)2  -3.14000    5.96657  -0.526
## factor(Sleep.Quality)Good:factor(Stress.Level)2  -0.60667    5.96657  -0.102
```

```
## factor(Sleep.Quality)Poor:factor(Stress.Level)2   2.14010    7.01910   0.305
## factor(Sleep.Quality)Fair:factor(Stress.Level)3   1.80000    7.13141   0.252
## factor(Sleep.Quality)Good:factor(Stress.Level)3   0.36667    6.17598   0.059
## factor(Sleep.Quality)Poor:factor(Stress.Level)3   1.89297    5.73767   0.330
## factor(Sleep.Quality)Fair:factor(Stress.Level)4  -4.25260    5.49512  -0.774
## factor(Sleep.Quality)Good:factor(Stress.Level)4  -4.11927    5.19786  -0.792
## factor(Sleep.Quality)Poor:factor(Stress.Level)4        NA         NA      NA
## factor(Sleep.Quality)Fair:factor(Stress.Level)5  -2.15000    6.67082  -0.322
## factor(Sleep.Quality)Good:factor(Stress.Level)5   1.68333    7.35089   0.229
## factor(Sleep.Quality)Poor:factor(Stress.Level)5   1.46086    6.04595   0.242
## factor(Sleep.Quality)Fair:factor(Stress.Level)6  -2.46667    6.04595  -0.408
## factor(Sleep.Quality)Good:factor(Stress.Level)6  -1.56667    6.30334  -0.249
## factor(Sleep.Quality)Poor:factor(Stress.Level)6  -2.00745    5.77710  -0.347
## factor(Sleep.Quality)Fair:factor(Stress.Level)7   0.06667    6.42817   0.010
## factor(Sleep.Quality)Good:factor(Stress.Level)7   3.33333    7.35089   0.453
## factor(Sleep.Quality)Poor:factor(Stress.Level)7  -0.37889    6.17598  -0.061
## factor(Sleep.Quality)Fair:factor(Stress.Level)8  -6.58104    7.35089  -0.895
## factor(Sleep.Quality)Good:factor(Stress.Level)8  -2.98419    7.35089  -0.406
## factor(Sleep.Quality)Poor:factor(Stress.Level)8  -9.31800    6.78890  -1.373
## factor(Sleep.Quality)Fair:factor(Stress.Level)9   4.63575    4.71699   0.983
## factor(Sleep.Quality)Good:factor(Stress.Level)9   5.04842    5.04267   1.001
## factor(Sleep.Quality)Poor:factor(Stress.Level)9        NA         NA      NA
##                                                  Pr(>|t|)
## (Intercept)                                      0.837357
## factor(Sleep.Quality)Fair                        0.692931
## factor(Sleep.Quality)Good                        0.848568
## factor(Sleep.Quality)Poor                        0.164499
## factor(Stress.Level)2                            0.812642
## factor(Stress.Level)3                            0.952740
## factor(Stress.Level)4                            0.276198
## factor(Stress.Level)5                            0.874216
## factor(Stress.Level)6                            0.759646
## factor(Stress.Level)7                            0.940611
## factor(Stress.Level)8                            0.461094
## factor(Stress.Level)9                            0.000191 ***
## factor(Sleep.Quality)Fair:factor(Stress.Level)2 0.600468
## factor(Sleep.Quality)Good:factor(Stress.Level)2 0.919321
## factor(Sleep.Quality)Poor:factor(Stress.Level)2 0.761404
## factor(Sleep.Quality)Fair:factor(Stress.Level)3 0.801513
## factor(Sleep.Quality)Good:factor(Stress.Level)3 0.952837
## factor(Sleep.Quality)Poor:factor(Stress.Level)3 0.742506
## factor(Sleep.Quality)Fair:factor(Stress.Level)4 0.441761
## factor(Sleep.Quality)Good:factor(Stress.Level)4 0.430913
## factor(Sleep.Quality)Poor:factor(Stress.Level)4       NA
## factor(Sleep.Quality)Fair:factor(Stress.Level)5 0.748245
## factor(Sleep.Quality)Good:factor(Stress.Level)5 0.819579
## factor(Sleep.Quality)Poor:factor(Stress.Level)5 0.809819
## factor(Sleep.Quality)Fair:factor(Stress.Level)6 0.684604
## factor(Sleep.Quality)Good:factor(Stress.Level)6 0.804484
## factor(Sleep.Quality)Poor:factor(Stress.Level)6 0.729334
## factor(Sleep.Quality)Fair:factor(Stress.Level)7 0.991757
## factor(Sleep.Quality)Good:factor(Stress.Level)7 0.651705
## factor(Sleep.Quality)Poor:factor(Stress.Level)7 0.951266
## factor(Sleep.Quality)Fair:factor(Stress.Level)8 0.373896
```

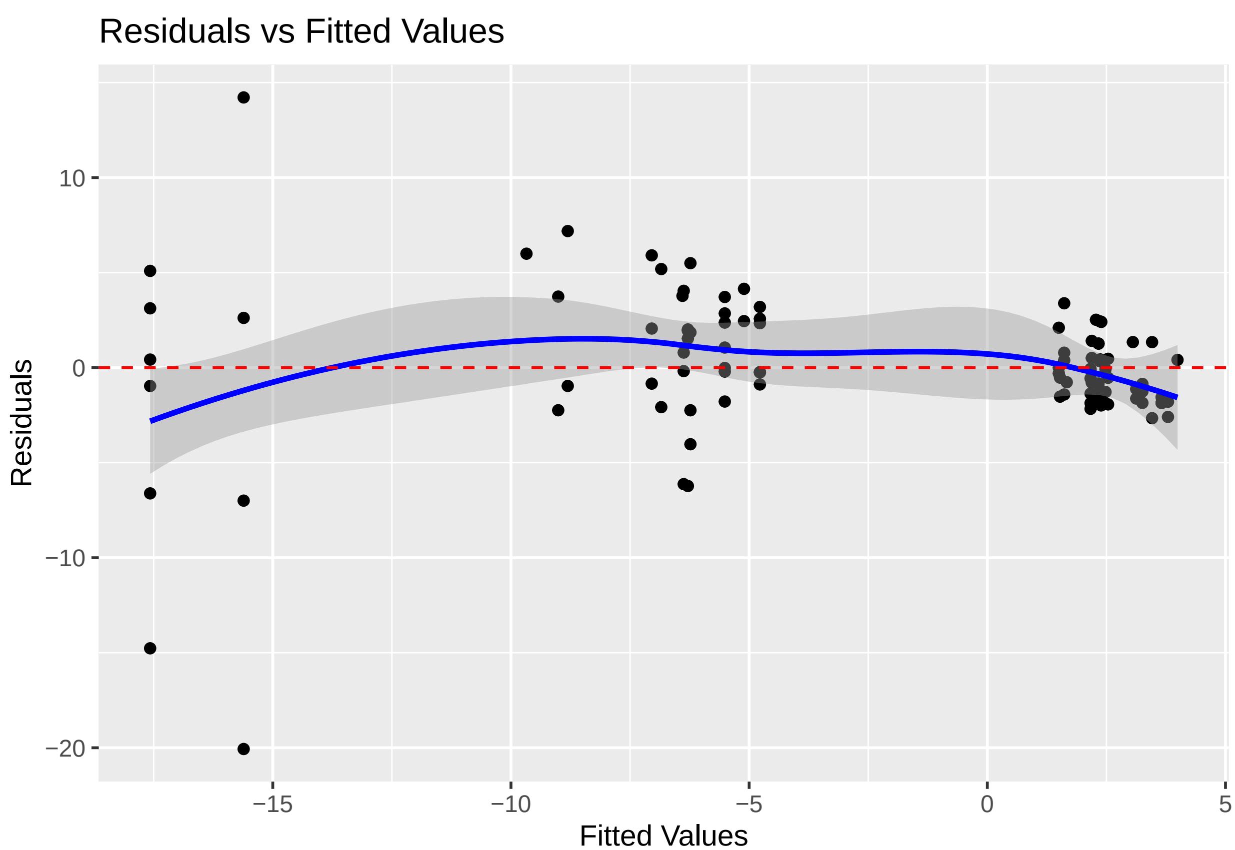
```
## factor(Sleep.Quality)Good:factor(Stress.Level)8 0.686083
## factor(Sleep.Quality)Poor:factor(Stress.Level)8 0.174544
## factor(Sleep.Quality)Fair:factor(Stress.Level)9 0.329306
## factor(Sleep.Quality)Good:factor(Stress.Level)9 0.320415
## factor(Sleep.Quality)Poor:factor(Stress.Level)9       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.367 on 66 degrees of freedom
## Multiple R-squared:  0.7705, Adjusted R-squared:  0.6558
## F-statistic: 6.716 on 33 and 66 DF,  p-value: 3.128e-11
```

Notice that this gives the same output as before. Similar to before we will stick with our refined model.

Now we will test for some of the assumptions in our model, using plots and stastical tests.

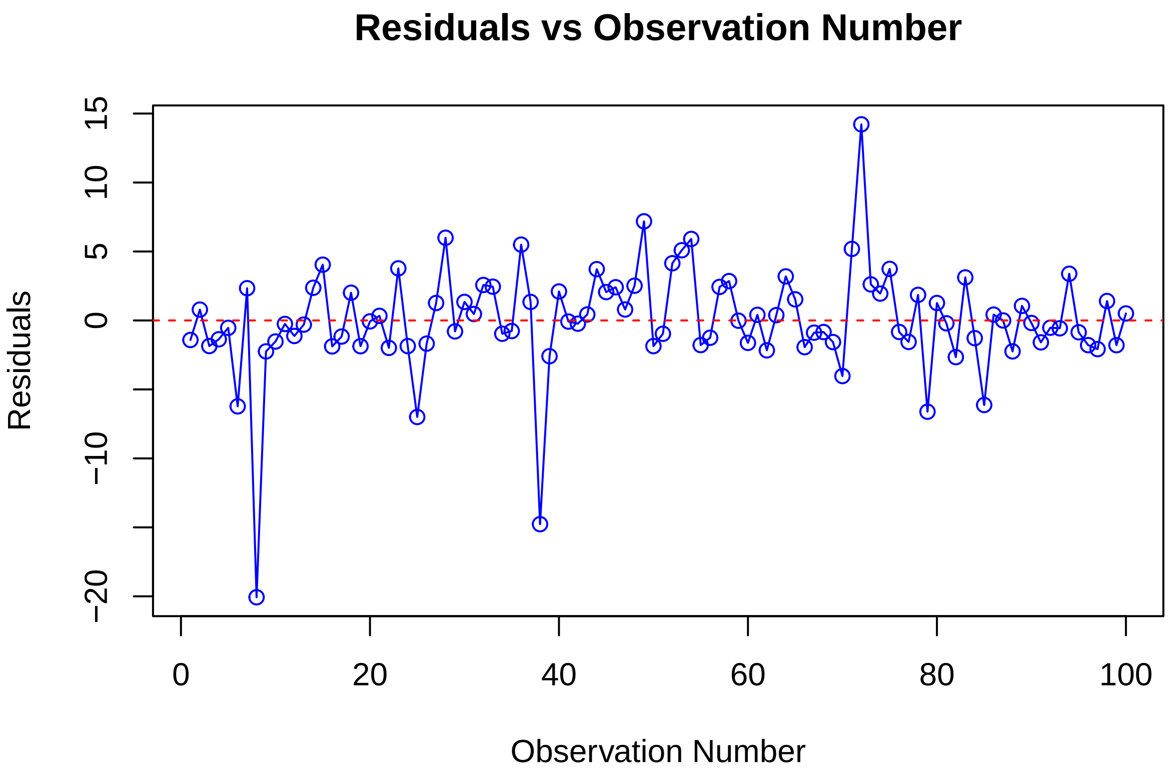
We can test for linearity by inspecting our residual plots.

```
## `geom_smooth()` using formula = 'y ~ x'
```



This plot suggests that the model does not quite pass the assumption of linearity. We can see that the residuals do not appear randomly scattered, rather they are clustered in 3 different groups, additionally we can see the presence of a curve in our line. Normally we could try transforming our data, but in this case, as mentioned before, we cannot add higher order terms. We also cannot perform a log transformation on our predictors or our responding variable, since our dummy variables contain 0 and our responding variable contains negative values.

To test our independence assumption, we can inspect the residuals in the figure below.

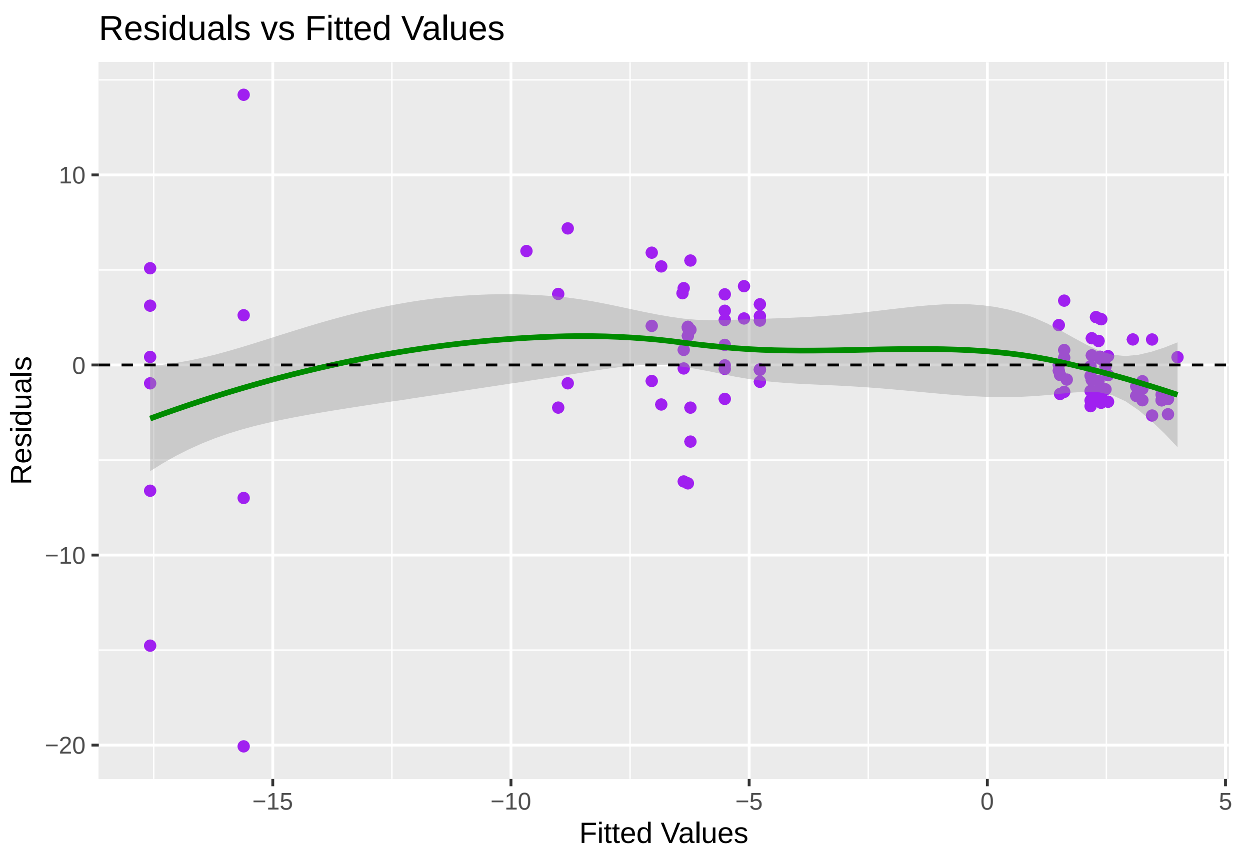


Given the randomly scattered point, and lack of trends, we can suggest that independence has most likely been met. Although there a few outliers, they do not appear to follow any pattern. Additionally, our responding variable is not considered time-series data.

To test for equal variance, we can inspect our residual plots and perform a Breusch-Pagan Test, where the hypothesis would be:

$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_n^2$ (heteroscedasticity is not present)H_a : at least one σ_i^2 is different from the others i = 1, 2, ..., n (het

```
## ‘geom_smooth()‘ using formula = ’y ~ x’
```

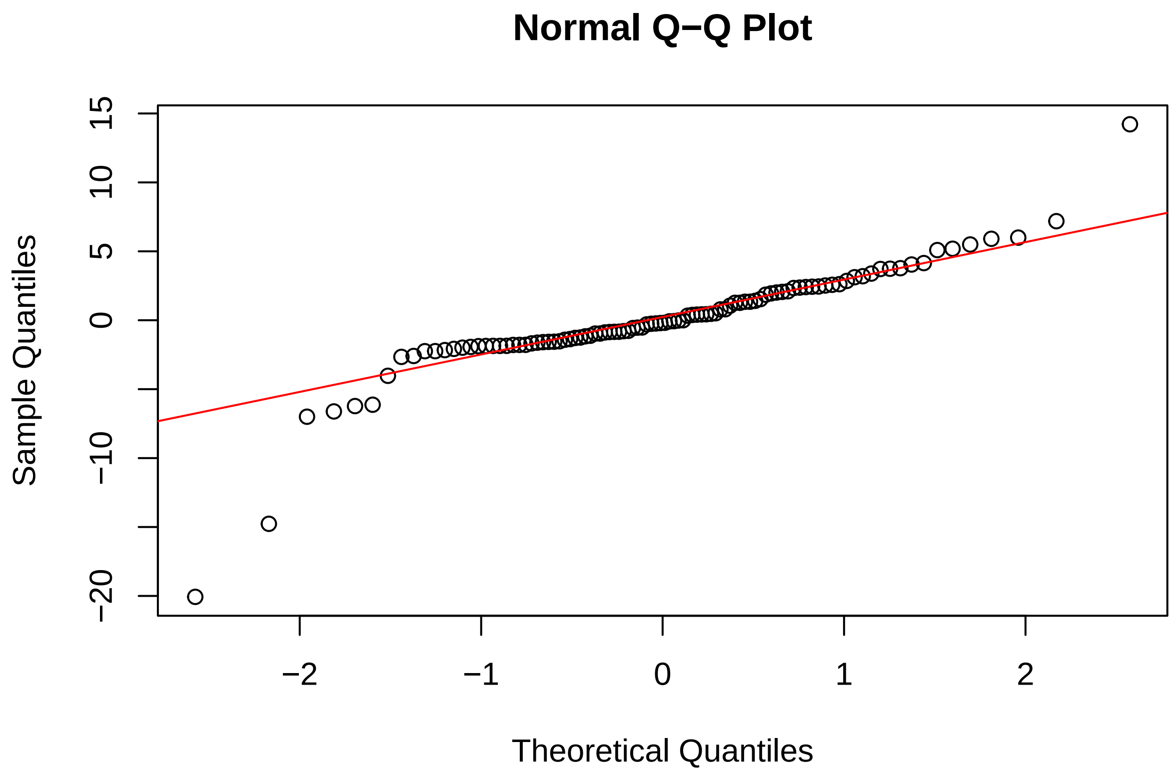



```
##
##  studentized Breusch-Pagan test
##
## data:  weight_interaction_model
## BP = 52.041, df = 33, p-value = 0.01872
```

As we can see the presence of patterns in our data points, and the line is curved, we believe heteroscedasticity is present. Our Breusch-Pagan test returned a p-value of 0.01872, which is less than 0.05. Therefore we reject the null hypothesis, and conclude there is statistically significant evidence of heteroscedasticity in the model.

To test our assumptions of normality, we can create our Q-Q plot and run a Shapiro-Wilks test, using the the hypothesis:

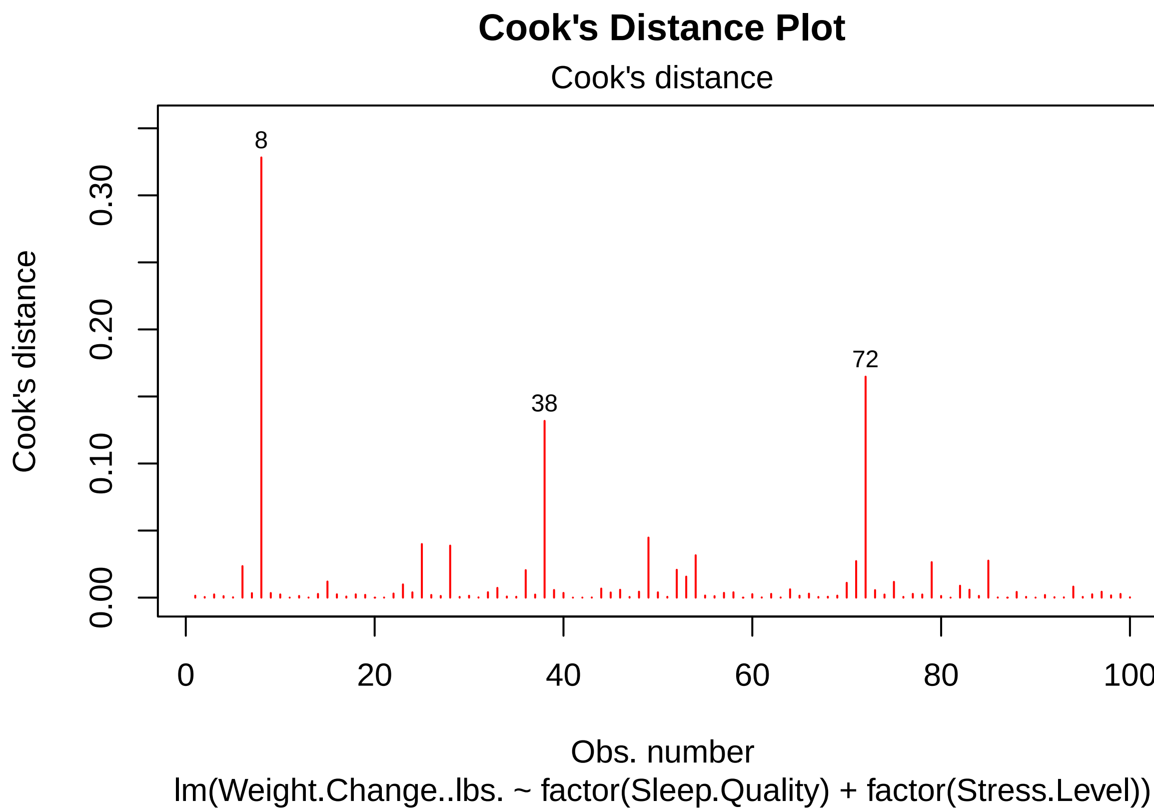
$$H_0 : \text{the residuals are normally distributed} H_a : \text{the residuals are not normally distributed}$$



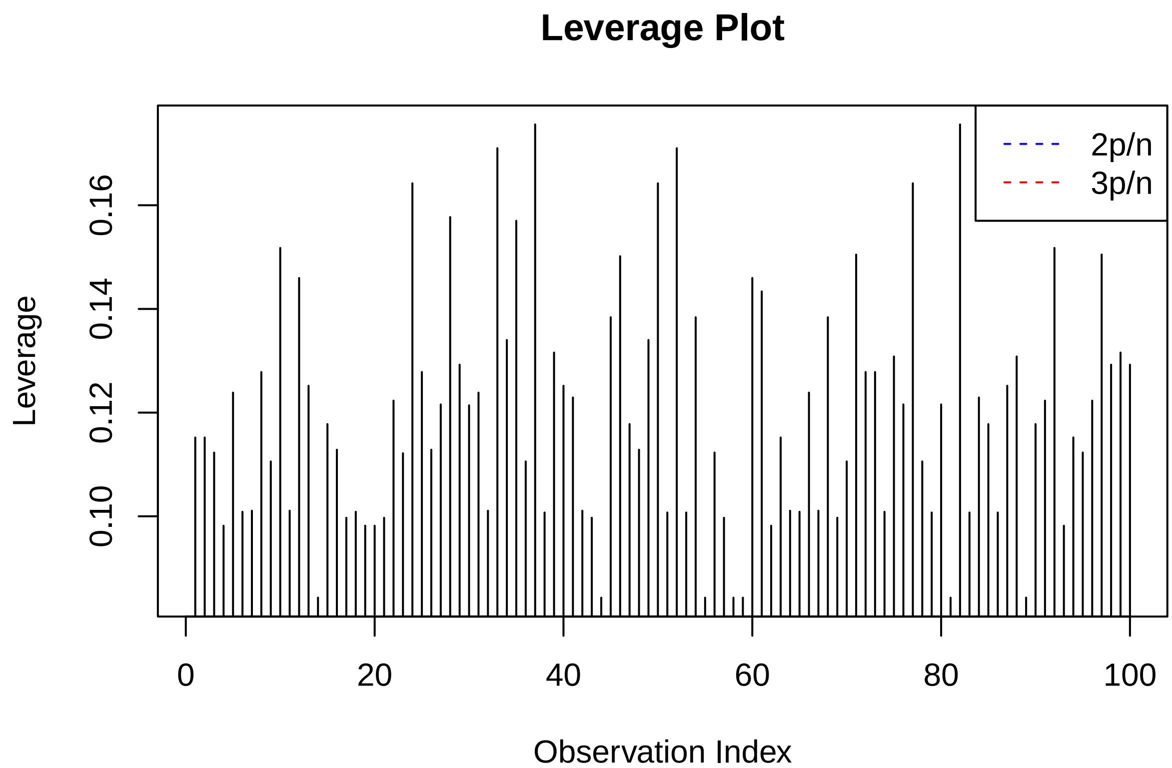

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(weight_refined_model)
## W = 0.83622, p-value = 3.793e-09
```

The Q-Q plot suggests that the data may not be perfectly normally distributed. The outliers might be indicative of non-normality. As our Shapiro-Wilks test returned a p-value of 3.793e-09, which is < 0.05, we reject the null hypothesis and conclude that our residuals are not normally distributed.

To check for outliers we can look at our Cook's Distance and Leverage plots.




```
##  [1] Participant.ID                Age
##  [3] Gender                        Current.Weight..lbs.
##  [5] BMR..Calories.                Daily.Calories.Consumed
##  [7] Daily.Caloric.Surplus.Deficit Weight.Change..lbs.
##  [9] Duration..weeks.              Physical.Activity.Level
## [11] Sleep.Quality                 Stress.Level
## [13] Final.Weight..lbs.
## <0 rows> (or 0-length row.names)
```



There are three influential points in the data, highlighted by the red vertical lines at observations 8, 38, and 72. Cook's Distance Values: These points have high Cook's distances, exceeding the threshold, suggesting they significantly influence the regression model.

High leverage: The outlier has an unusual combination of predictor values, making it influential. High residual: The outlier has a large difference between its observed and predicted values, suggesting it's not well explained by the model.

The data points at observations 8, 38, and 72 are influential because they significantly affect the estimated coefficients of the regression model.

Although all of our assumptions were not met, we will still present our final model.

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ factor(Sleep.Quality) + factor(Stress.Level),
##     data = weight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.0668  -1.5978  -0.1951   2.0673  14.2170
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  1.66777    1.64232   1.015    0.313
## factor(Sleep.Quality)Fair    0.66673    1.37884   0.484    0.630
## factor(Sleep.Quality)Good    0.86678    1.41286   0.613    0.541
## factor(Sleep.Quality)Poor   -7.89997    1.28104  -6.167 2.06e-08 ***
```

```
## factor(Stress.Level)2      -0.16713    1.69144  -0.099    0.922
## factor(Stress.Level)3       0.72075    1.68828   0.427    0.670
## factor(Stress.Level)4       1.12429    2.01231   0.559    0.578
## factor(Stress.Level)5       1.45804    1.78473   0.817    0.416
## factor(Stress.Level)6      -0.05374    1.71822  -0.031    0.975
## factor(Stress.Level)7      -0.14142    1.82197  -0.078    0.938
## factor(Stress.Level)8     -11.34225    1.77596  -6.387 7.78e-09 ***
## factor(Stress.Level)9      -9.37917    1.86760  -5.022 2.65e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.145 on 88 degrees of freedom
## Multiple R-squared:  0.7244, Adjusted R-squared:   0.69
## F-statistic: 21.03 on 11 and 88 DF,  p-value: < 2.2e-16
```

$$\text{Weight.Change (lbs)} = 1.66777 + 0.66673 \cdot \text{SleepQualityFair} + 0.86678 \cdot \text{SleepQualityGood} - 7.89997 \cdot \text{SleepQualityPoor} - 0.16713 \cdot \text{S}$$

In conclusion the refined model explores the relationship between weight change and two key factors: sleep quality and stress level.

Now that we have our final model, we can begin interpreting it.

Intercept: When sleep quality is excellent, and our stress level is 1, we can expect weight change to increase by 1.66777 pounds.

Sleep Quality:

- Participants with fair sleep quality gain 0.66673 pounds more weight on average compared to those with excellent sleep quality.
- Participants with good sleep quality gain 0.86678 pounds more weight on average compared to those with excellent sleep quality.
- Participants with poor sleep quality lose 7.89997 pounds more weight on average compared to those with excellent sleep quality.

Stress Level:

- Participants with a stress level of 2 lose 0.16713 pounds more weight on average compared to those with a stress level of 1.
- Participants with a stress level of 3 gain 0.72075 pounds more weight on average compared to those with a stress level of 1.
- Participants with a stress level of 4 gain 1.12429 pounds more weight on average compared to those with a stress level of 1.
- Participants with a stress level of 5 gain 1.45804 pounds more weight on average compared to those with a stress level of 1.
- Participants with a stress level of 6 lose 0.05374 pounds more weight on average compared to those with a stress level of 1.
- Participants with a stress level of 7 lose 0.14142 pounds more weight on average compared to those with a stress level of 1.
- Participants with a stress level of 8 lose 11.34225 pounds more weight on average compared to those with a stress level of 1.
- Participants with a stress level of 9 lose 9.37917 pounds more weight on average compared to those with a stress level of 1.

These findings suggest that worse sleep quality and higher stress levels are associated with greater weight loss.

Overall, the model explains a moderate proportion of variance in weight change (Adjusted R-squared = 0.69), indicating that these two factors contribute to understanding weight changes, but other un-explored factors might also play a significant role.

3.3 Model Including All Variables

To determine the best model for predicting weight change, given our dataset, we begin by inspecting our dataset.

```
##   Participant.ID Age Gender Current.Weight..lbs. BMR..Calories.
## 1              1  56      M                228.4         3102.3
## 2              2  46      F                165.4         2275.5
## 3              3  32      F                142.8         2119.4
## 4              4  25      F                145.5         2181.3
## 5              5  38      M                155.5         2463.8
## 6              6  56      F                152.9         2100.6
##   Daily.Calories.Consumed Daily.Caloric.Surplus.Deficit Weight.Change..lbs.
## 1                  3916.0                         813.7              0.2000
## 2                  3823.0                        1547.5              2.4000
## 3                  2785.4                         666.0              1.4000
## 4                  2587.3                         406.0              0.8000
## 5                  3312.8                         849.0              2.0000
## 6                  2262.4                         161.9            -12.5135
##   Duration..weeks. Physical.Activity.Level Sleep.Quality Stress.Level
## 1                1               Sedentary     Excellent            6
## 2                6             Very Active     Excellent            6
## 3                7               Sedentary          Good            3
## 4                8               Sedentary          Fair            2
## 5               10          Lightly Active          Good            1
## 6                9               Sedentary          Poor            6
##   Final.Weight..lbs.
## 1              228.6
## 2              167.8
## 3              144.2
## 4              146.3
## 5              157.5
## 6              140.4
```

Our first step is to build our full model, and determine whether any assumptions have been broken. Do note, that we chose to remove the variables "Current.Weight..lbs." and "Final.Weight..lbs." because our responding variable, "Weight.Change..lbs." is just a result of subtracting the two variables, making weight change dependent on both current and final weight. To ensure that at least one of our predictors is significant, we test the hypothesis:

$H_0 : \beta_1 = \beta_2 = ... = \beta_i = 0$ (i=1,2,...,p)
$H_a :$ At least one $\beta_i \neq 0$ (i=1,2,...,p)

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ Age + factor(Gender) + BMR..Calories. +
##     Daily.Calories.Consumed + Daily.Caloric.Surplus.Deficit +
```

```
##     Duration..weeks. + factor(Physical.Activity.Level) + factor(Sleep.Quality) +
##     factor(Stress.Level), data = weight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.5881  -1.4426   0.2319   2.1582   9.7510
##
## Coefficients:
##                                                Estimate Std. Error t value
## (Intercept)                                    7.032845   4.940503   1.424
## Age                                           -0.002834   0.036018  -0.079
## factor(Gender)M                                0.263772   1.150845   0.229
## BMR..Calories.                                -6.915519  10.273016  -0.673
## Daily.Calories.Consumed                        6.913309  10.272867   0.673
## Daily.Caloric.Surplus.Deficit                 -6.911143  10.273082  -0.673
## Duration..weeks.                              -0.395738   0.122177  -3.239
## factor(Physical.Activity.Level)Moderately Active  -0.316972   1.247677  -0.254
## factor(Physical.Activity.Level)Sedentary       0.007914   1.327862   0.006
## factor(Physical.Activity.Level)Very Active     0.751174   1.520393   0.494
## factor(Sleep.Quality)Fair                      1.544141   1.366646   1.130
## factor(Sleep.Quality)Good                      2.109521   1.416478   1.489
## factor(Sleep.Quality)Poor                     -7.411684   1.271290  -5.830
## factor(Stress.Level)2                         -0.153266   1.717603  -0.089
## factor(Stress.Level)3                          1.489659   1.714714   0.869
## factor(Stress.Level)4                          1.006707   1.974665   0.510
## factor(Stress.Level)5                          0.770360   1.798465   0.428
## factor(Stress.Level)6                         -0.849362   1.690039  -0.503
## factor(Stress.Level)7                          0.727956   1.848089   0.394
## factor(Stress.Level)8                        -11.058953   1.754401  -6.304
## factor(Stress.Level)9                         -9.822984   1.838339  -5.343
##                                               Pr(>|t|)
## (Intercept)                                    0.15853
## Age                                            0.93748
## factor(Gender)M                                0.81931
## BMR..Calories.                                 0.50280
## Daily.Calories.Consumed                        0.50293
## Daily.Caloric.Surplus.Deficit                  0.50307
## Duration..weeks.                               0.00176 **
## factor(Physical.Activity.Level)Moderately Active  0.80012
## factor(Physical.Activity.Level)Sedentary       0.99526
## factor(Physical.Activity.Level)Very Active     0.62263
## factor(Sleep.Quality)Fair                      0.26195
## factor(Sleep.Quality)Good                      0.14040
## factor(Sleep.Quality)Poor                     1.15e-07 ***
## factor(Stress.Level)2                          0.92912
## factor(Stress.Level)3                          0.38762
## factor(Stress.Level)4                          0.61160
## factor(Stress.Level)5                          0.66957
## factor(Stress.Level)6                          0.61667
## factor(Stress.Level)7                          0.69472
## factor(Stress.Level)8                         1.55e-08 ***
## factor(Stress.Level)9                         8.58e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 3.925 on 79 degrees of freedom
## Multiple R-squared:  0.7781, Adjusted R-squared:  0.7219
## F-statistic: 13.85 on 20 and 79 DF,  p-value: < 2.2e-16
```

Based on our output, we can see that at least one predictor has a p-value < 0.05, therefore we reject our null hypothesis and conclude that at least one predictor significantly affects weight change. Before determining which predictors to include in our model, we want to test for multicollinearity. This can be done using the VIF method.

```
##                                         GVIF Df GVIF^(1/(2*Df))
## Age                             1.244553e+00  1        1.115595
## factor(Gender)                  2.106715e+00  1        1.451453
## BMR..Calories.                  9.005157e+07  1     9489.550825
## Daily.Calories.Consumed         1.786538e+08  1    13366.144669
## Daily.Caloric.Surplus.Deficit   9.361072e+07  1     9675.263124
## Duration..weeks.                1.185115e+00  1        1.088630
## factor(Physical.Activity.Level) 4.324149e+00  3        1.276390
## factor(Sleep.Quality)           1.580526e+00  3        1.079279
## factor(Stress.Level)            3.048675e+00  8        1.072153
```

```
##
## Call:
## imcdiag(mod = weight_model_full, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                                                       VIF detection
## Age                                          1.244600e+00         0
## factor(Gender)M                              2.106700e+00         0
## BMR..Calories.                               9.005157e+07         1
## Daily.Calories.Consumed                      1.786538e+08         1
## Daily.Caloric.Surplus.Deficit                9.361071e+07         1
## Duration..weeks.                             1.185100e+00         0
## factor(Physical.Activity.Level)Moderately Active 1.894200e+00         0
## factor(Physical.Activity.Level)Sedentary     1.898400e+00         0
## factor(Physical.Activity.Level)Very Active   2.656800e+00         0
## factor(Sleep.Quality)Fair                    2.210900e+00         0
## factor(Sleep.Quality)Good                    2.234400e+00         0
## factor(Sleep.Quality)Poor                    2.471100e+00         0
## factor(Stress.Level)2                        2.305200e+00         0
## factor(Stress.Level)3                        2.297400e+00         0
## factor(Stress.Level)4                        1.647400e+00         0
## factor(Stress.Level)5                        2.055000e+00         0
## factor(Stress.Level)6                        2.096500e+00         0
## factor(Stress.Level)7                        1.994900e+00         0
## factor(Stress.Level)8                        1.955600e+00         0
## factor(Stress.Level)9                        1.796200e+00         0
##
## Multicollinearity may be due to BMR..Calories. Daily.Calories.Consumed Daily.Caloric.Surplus.Deficit
##
## 1 --> COLLINEARITY is detected by the test
```

```
## 0 --> COLLINEARITY is not detected by the test
##
## ===================================
```

From our output, we can see that three variables have a VIF score > 5, suggesting multicollinearity. These variables are "BMR..Calories.", "Daily.Calories.Consumed", and "Daily.Caloric.Surplus.Deficit". Based on the definitions provided in the data set, "Daily.Caloric.Surplus.Deficit" is the difference between calories consumed and BMR, therefore we chose to keep "Daily.Caloric.Surplus.Deficit" and remove "BMR..Calories.", and "Daily.Calories.Consumed".

Following the removal of those variables, we will determine which predictors are significant using the hypothesis test:

$H_0 : \beta_i = 0$ (i=1,2,...,p)
$H_a : \beta_i \neq 0$ (i=1,2,...,p)

This can be achieved by looking at the individual t-tests from the summary of our adjusted model.

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ Daily.Caloric.Surplus.Deficit +
##     Age + factor(Gender) + Duration..weeks. + factor(Physical.Activity.Level) +
##     factor(Sleep.Quality) + factor(Stress.Level), data = weight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.221  -1.637   0.325   2.286   9.717
##
## Coefficients:
##                                                  Estimate Std. Error t value
## (Intercept)                                      1.098664   3.061275   0.359
## Daily.Caloric.Surplus.Deficit                    0.002679   0.001871   1.432
## Age                                              0.003592   0.035076   0.102
## factor(Gender)M                                 -0.768568   0.922416  -0.833
## Duration..weeks.                                -0.402046   0.121077  -3.321
## factor(Physical.Activity.Level)Moderately Active  -0.618016   1.234619  -0.501
## factor(Physical.Activity.Level)Sedentary        -0.016837   1.330946  -0.013
## factor(Physical.Activity.Level)Very Active       0.360470   1.505470   0.239
## factor(Sleep.Quality)Fair                        1.587961   1.365649   1.163
## factor(Sleep.Quality)Good                        2.256215   1.418302   1.591
## factor(Sleep.Quality)Poor                       -7.353535   1.273937  -5.772
## factor(Stress.Level)2                            0.382922   1.680623   0.228
## factor(Stress.Level)3                            1.499079   1.707093   0.878
## factor(Stress.Level)4                            1.272710   1.961313   0.649
## factor(Stress.Level)5                            1.420505   1.757602   0.808
## factor(Stress.Level)6                           -0.479314   1.665477  -0.288
## factor(Stress.Level)7                            0.924742   1.821196   0.508
## factor(Stress.Level)8                          -10.782512   1.735068  -6.214
## factor(Stress.Level)9                           -9.275712   1.804889  -5.139
##                                                Pr(>|t|)
## (Intercept)                                     0.72061
## Daily.Caloric.Surplus.Deficit                   0.15610
## Age                                             0.91868
## factor(Gender)M                                 0.40718
## Duration..weeks.                                0.00135 **
```

```
## factor(Physical.Activity.Level)Moderately Active  0.61803
## factor(Physical.Activity.Level)Sedentary          0.98994
## factor(Physical.Activity.Level)Very Active        0.81137
## factor(Sleep.Quality)Fair                         0.24833
## factor(Sleep.Quality)Good                         0.11555
## factor(Sleep.Quality)Poor                        1.39e-07 ***
## factor(Stress.Level)2                             0.82034
## factor(Stress.Level)3                             0.38246
## factor(Stress.Level)4                             0.51823
## factor(Stress.Level)5                             0.42134
## factor(Stress.Level)6                             0.77424
## factor(Stress.Level)7                             0.61300
## factor(Stress.Level)8                            2.11e-08 ***
## factor(Stress.Level)9                            1.87e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.939 on 81 degrees of freedom
## Multiple R-squared:  0.7709, Adjusted R-squared:   0.72
## F-statistic: 15.14 on 18 and 81 DF,  p-value: < 2.2e-16
```

Based on this output, we can see there are 4 instances where the null hypothesis is rejected (p-value < 0.05), implying that these variables are significant and should be included in our model. These variables are:

- Duration..weeks.
- factor(Sleep.Quality)Poor
- factor(Stress.Level)8
- factor(Stress.Level)9

Despite the fact that only some of the dummies are significant, for the categories "Sleep.Quality" and "Stress.Level", we must include the full variable in our model. Therefore our model will include the predictors Sleep.Quality, Stress.Level and Duration..weeks.

NOTE An attempt was made to use stepwise regression to determine which variables were best to keep, with the code for that being included in the appendix. Upon running the stepwise, it was determined that a 4th variable, Daily,Caloric.Surplus.Deficit should be included. In the end we chose not to include this variable for a couple reasons. First, our categorical variables had many levels, in which case it is often better to just manually select significant predictors using t-tests from our summary table. When performing this manual selection, we found the variable the stepwise selection wanted to include, had a p-value 0f 0.15610, and a coefficient of 0.002679. This implied the variable had no significant effect. Based on this, and to avoid overfitting and complicating our model, we chose to not include this variable

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ factor(Sleep.Quality) + Duration..weeks. +
##     factor(Stress.Level), data = weight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.2553  -1.9040   0.2041   2.3282  12.3789
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  3.7718     1.7173   2.196   0.0307 *
```

```
## factor(Sleep.Quality)Fair   1.4083     1.3413   1.050   0.2967
## factor(Sleep.Quality)Good   1.8466     1.3895   1.329   0.1873
## factor(Sleep.Quality)Poor  -7.4041     1.2362  -5.990 4.62e-08 ***
## Duration..weeks.           -0.3650     0.1204  -3.032   0.0032 **
## factor(Stress.Level)2      -0.3405     1.6188  -0.210   0.8339
## factor(Stress.Level)3       0.4781     1.6168   0.296   0.7682
## factor(Stress.Level)4       0.6451     1.9312   0.334   0.7391
## factor(Stress.Level)5       0.9831     1.7142   0.573   0.5678
## factor(Stress.Level)6      -0.3863     1.6471  -0.235   0.8151
## factor(Stress.Level)7       0.4839     1.7548   0.276   0.7834
## factor(Stress.Level)8     -11.3423     1.6987  -6.677 2.19e-09 ***
## factor(Stress.Level)9      -9.7760     1.7911  -5.458 4.48e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.964 on 87 degrees of freedom
## Multiple R-squared:  0.7507, Adjusted R-squared:  0.7164
## F-statistic: 21.84 on 12 and 87 DF,  p-value: < 2.2e-16
```

We can see that all predictors are now significant. Before moving on to interaction terms, we should check again for multicollinearity, just to be certain:

```
##                          GVIF Df GVIF^(1/(2*Df))
## factor(Sleep.Quality) 1.303780  3        1.045203
## Duration..weeks.      1.128093  1        1.062117
## factor(Stress.Level)  1.308020  8        1.016924
```

```
##
## Call:
## imcdiag(mod = weight_model_adj2, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                               VIF detection
## factor(Sleep.Quality)Fair 2.0880         0
## factor(Sleep.Quality)Good 2.1080         0
## factor(Sleep.Quality)Poor 2.2907         0
## Duration..weeks.          1.1281         0
## factor(Stress.Level)2     2.0076         0
## factor(Stress.Level)3     2.0025         0
## factor(Stress.Level)4     1.5448         0
## factor(Stress.Level)5     1.8305         0
## factor(Stress.Level)6     1.9523         0
## factor(Stress.Level)7     1.7634         0
## factor(Stress.Level)8     1.7974         0
## factor(Stress.Level)9     1.6718         0
##
## NOTE:  VIF Method Failed to detect multicollinearity
##
##
## 0 --> COLLINEARITY is not detected by the test
##
## ===================================
```

As we can see from our output, there are no issues with multicollinearity.

Now we can begin checking for interaction terms. We can start by testing all interaction terms for our model using the hypothesis test:

$H_0 : \beta_{interaction} = 0$ (i=1,2,...,p)
$H_a : \beta_{interaction} \neq 0$ (i=1,2,...,p)

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ (factor(Sleep.Quality) + Duration..weeks. +
##     factor(Stress.Level))^2, data = weight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.0955  -0.5464   0.0000   0.9004   9.1887
##
## Coefficients: (2 not defined because of singularities)
##                                              Estimate Std. Error t value
## (Intercept)                                  0.173136   3.385089   0.051
## factor(Sleep.Quality)Fair                    0.630967   4.305253   0.147
## factor(Sleep.Quality)Good                   -0.515159   4.272899  -0.121
## factor(Sleep.Quality)Poor                    1.111739   4.146862   0.268
## Duration..weeks.                             0.242288   0.465477   0.521
## factor(Stress.Level)2                        0.403610   4.337921   0.093
## factor(Stress.Level)3                       -1.956105   3.862834  -0.506
## factor(Stress.Level)4                       -1.834648   3.855770  -0.476
## factor(Stress.Level)5                       -4.257771   4.621632  -0.921
## factor(Stress.Level)6                        0.391914   3.849036   0.102
## factor(Stress.Level)7                       -2.320497   4.601681  -0.504
## factor(Stress.Level)8                        1.171047   5.307511   0.221
## factor(Stress.Level)9                       -3.418483   3.543061  -0.965
## factor(Sleep.Quality)Fair:Duration..weeks.  -0.009411   0.393084  -0.024
## factor(Sleep.Quality)Good:Duration..weeks.   0.045802   0.439363   0.104
## factor(Sleep.Quality)Poor:Duration..weeks.  -1.208733   0.442719  -2.730
## factor(Sleep.Quality)Fair:factor(Stress.Level)2 -1.610491   4.861653  -0.331
## factor(Sleep.Quality)Good:factor(Stress.Level)2  0.638105   4.726767   0.135
## factor(Sleep.Quality)Poor:factor(Stress.Level)2 -0.271656   5.459211  -0.050
## factor(Sleep.Quality)Fair:factor(Stress.Level)3 -2.208279   6.084939  -0.363
## factor(Sleep.Quality)Good:factor(Stress.Level)3 -3.121900   5.144181  -0.607
## factor(Sleep.Quality)Poor:factor(Stress.Level)3 -0.806757   4.378219  -0.184
## factor(Sleep.Quality)Fair:factor(Stress.Level)4 -1.857250   7.660308  -0.242
## factor(Sleep.Quality)Good:factor(Stress.Level)4 -0.499270   5.817921  -0.086
## factor(Sleep.Quality)Poor:factor(Stress.Level)4        NA         NA      NA
## factor(Sleep.Quality)Fair:factor(Stress.Level)5 -0.155156   5.133104  -0.030
## factor(Sleep.Quality)Good:factor(Stress.Level)5 -4.340225   6.172293  -0.703
## factor(Sleep.Quality)Poor:factor(Stress.Level)5 -0.442804   4.533601  -0.098
## factor(Sleep.Quality)Fair:factor(Stress.Level)6 -0.991839   4.720897  -0.210
## factor(Sleep.Quality)Good:factor(Stress.Level)6 -2.044183   5.054225  -0.404
## factor(Sleep.Quality)Poor:factor(Stress.Level)6 -3.334619   4.372421  -0.763
## factor(Sleep.Quality)Fair:factor(Stress.Level)7  0.052776   5.351148   0.010
## factor(Sleep.Quality)Good:factor(Stress.Level)7  1.618175   6.488306   0.249
## factor(Sleep.Quality)Poor:factor(Stress.Level)7  1.392698   5.095646   0.273
## factor(Sleep.Quality)Fair:factor(Stress.Level)8 -4.308983   5.585545  -0.771
## factor(Sleep.Quality)Good:factor(Stress.Level)8 -3.359882   5.572127  -0.603
```

```
## factor(Sleep.Quality)Poor:factor(Stress.Level)8 -4.244128   5.215951  -0.814
## factor(Sleep.Quality)Fair:factor(Stress.Level)9  3.681779   3.459445   1.064
## factor(Sleep.Quality)Good:factor(Stress.Level)9  4.692112   3.573977   1.313
## factor(Sleep.Quality)Poor:factor(Stress.Level)9        NA         NA      NA
## Duration..weeks.:factor(Stress.Level)2          -0.043603   0.388747  -0.112
## Duration..weeks.:factor(Stress.Level)3           0.472603   0.462794   1.021
## Duration..weeks.:factor(Stress.Level)4           0.335902   0.851910   0.394
## Duration..weeks.:factor(Stress.Level)5           0.924639   0.550379   1.680
## Duration..weeks.:factor(Stress.Level)6           0.091339   0.400593   0.228
## Duration..weeks.:factor(Stress.Level)7           0.198939   0.543024   0.366
## Duration..weeks.:factor(Stress.Level)8          -1.079441   0.518387  -2.082
## Duration..weeks.:factor(Stress.Level)9          -1.245371   0.377083  -3.303
##                                                 Pr(>|t|)
## (Intercept)                                      0.95940
## factor(Sleep.Quality)Fair                        0.88403
## factor(Sleep.Quality)Good                        0.90448
## factor(Sleep.Quality)Poor                        0.78965
## Duration..weeks.                                 0.60483
## factor(Stress.Level)2                            0.92621
## factor(Stress.Level)3                            0.61464
## factor(Stress.Level)4                            0.63612
## factor(Stress.Level)5                            0.36101
## factor(Stress.Level)6                            0.91928
## factor(Stress.Level)7                            0.61612
## factor(Stress.Level)8                            0.82621
## factor(Stress.Level)9                            0.33893
## factor(Sleep.Quality)Fair:Duration..weeks.       0.98099
## factor(Sleep.Quality)Good:Duration..weeks.       0.91736
## factor(Sleep.Quality)Poor:Duration..weeks.       0.00853 **
## factor(Sleep.Quality)Fair:factor(Stress.Level)2  0.74173
## factor(Sleep.Quality)Good:factor(Stress.Level)2  0.89312
## factor(Sleep.Quality)Poor:factor(Stress.Level)2  0.96050
## factor(Sleep.Quality)Fair:factor(Stress.Level)3  0.71809
## factor(Sleep.Quality)Good:factor(Stress.Level)3  0.54647
## factor(Sleep.Quality)Poor:factor(Stress.Level)3  0.85450
## factor(Sleep.Quality)Fair:factor(Stress.Level)4  0.80935
## factor(Sleep.Quality)Good:factor(Stress.Level)4  0.93193
## factor(Sleep.Quality)Poor:factor(Stress.Level)4       NA
## factor(Sleep.Quality)Fair:factor(Stress.Level)5  0.97600
## factor(Sleep.Quality)Good:factor(Stress.Level)5  0.48497
## factor(Sleep.Quality)Poor:factor(Stress.Level)5  0.92255
## factor(Sleep.Quality)Fair:factor(Stress.Level)6  0.83438
## factor(Sleep.Quality)Good:factor(Stress.Level)6  0.68748
## factor(Sleep.Quality)Poor:factor(Stress.Level)6  0.44899
## factor(Sleep.Quality)Fair:factor(Stress.Level)7  0.99217
## factor(Sleep.Quality)Good:factor(Stress.Level)7  0.80400
## factor(Sleep.Quality)Poor:factor(Stress.Level)7  0.78566
## factor(Sleep.Quality)Fair:factor(Stress.Level)8  0.44380
## factor(Sleep.Quality)Good:factor(Stress.Level)8  0.54905
## factor(Sleep.Quality)Poor:factor(Stress.Level)8  0.41940
## factor(Sleep.Quality)Fair:factor(Stress.Level)9  0.29194
## factor(Sleep.Quality)Good:factor(Stress.Level)9  0.19478
## factor(Sleep.Quality)Poor:factor(Stress.Level)9       NA
## Duration..weeks.:factor(Stress.Level)2           0.91111
```

```
## Duration..weeks.:factor(Stress.Level)3          0.31172
## Duration..weeks.:factor(Stress.Level)4          0.69492
## Duration..weeks.:factor(Stress.Level)5          0.09873 .
## Duration..weeks.:factor(Stress.Level)6          0.82050
## Duration..weeks.:factor(Stress.Level)7          0.71553
## Duration..weeks.:factor(Stress.Level)8          0.04207 *
## Duration..weeks.:factor(Stress.Level)9          0.00170 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.084 on 54 degrees of freedom
## Multiple R-squared:  0.9064, Adjusted R-squared:  0.8284
## F-statistic: 11.62 on 45 and 54 DF,  p-value: 7.362e-16
```

Based on our output, we can see the significant variables (with a p-value <0.05) are "factor(Sleep.Quality)Poor:Duration..weeks "Duration..weeks.:factor(Stress.Level)8", "Duration..weeks.:factor(Stress.Level)9". Although none of the initial predictors are considered significant, they are all parts of significant interaction terms, and must therefore be included in the model, therefore our interaction model would include 'Sleep.Quality', 'Duration..weeks.', 'Stress.Level', 'Sleep.Quality':'Duration..weeks.', and 'Duration..weeks.':'Stress.Level'.

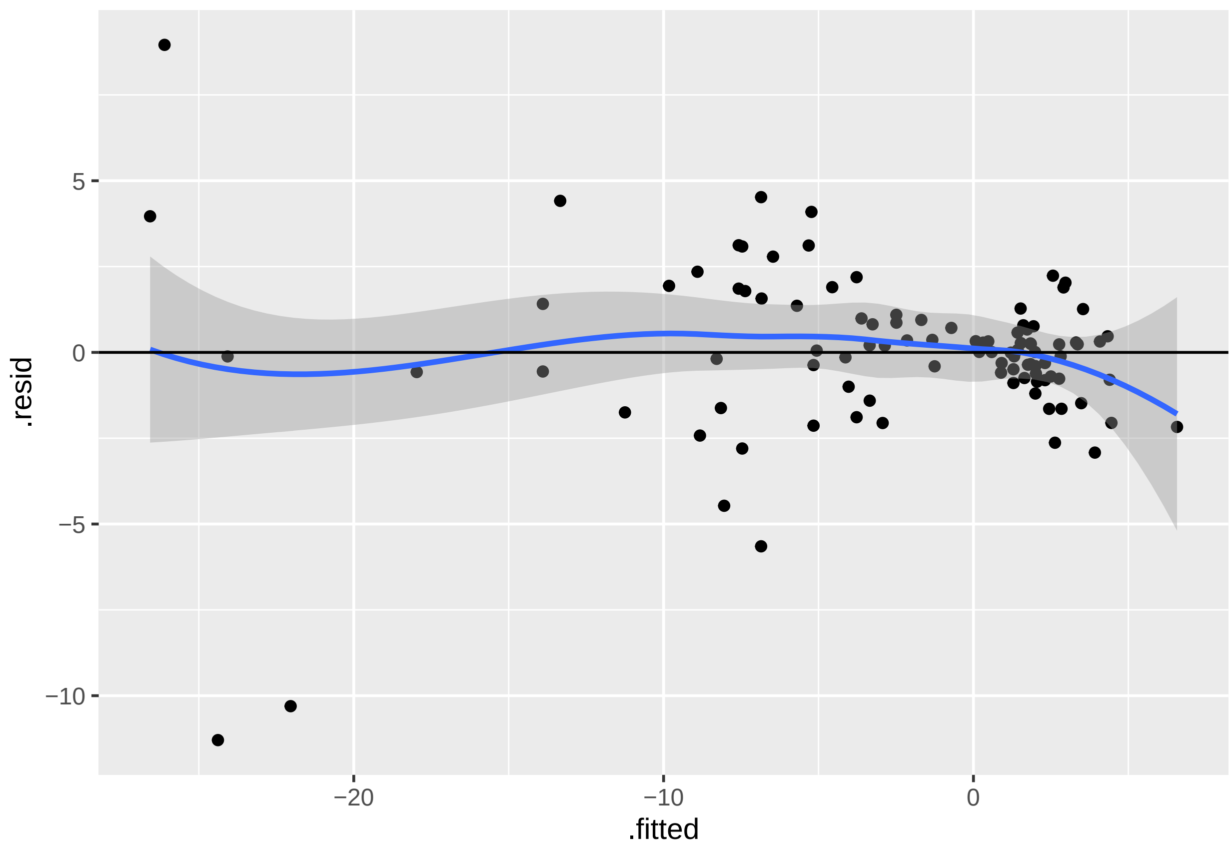
```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ factor(Sleep.Quality) + Duration..weeks. +
##     factor(Stress.Level) + factor(Sleep.Quality) * Duration..weeks. +
##     Duration..weeks. * factor(Stress.Level), data = weight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.2934  -0.7740   0.0613   0.9565   8.9577
##
## Coefficients:
##                                             Estimate Std. Error t value
## (Intercept)                                 0.846572   2.428365   0.349
## factor(Sleep.Quality)Fair                  -0.060956   2.125053  -0.029
## factor(Sleep.Quality)Good                  -0.731905   2.427593  -0.301
## factor(Sleep.Quality)Poor                  -0.876532   2.005304  -0.437
## Duration..weeks.                            0.266336   0.345765   0.770
## factor(Stress.Level)2                      -0.053039   2.626820  -0.020
## factor(Stress.Level)3                      -1.509118   2.582405  -0.584
## factor(Stress.Level)4                      -0.529393   3.095647  -0.171
## factor(Stress.Level)5                      -2.710051   2.845423  -0.952
## factor(Stress.Level)6                      -0.969032   2.490558  -0.389
## factor(Stress.Level)7                      -2.717472   3.578434  -0.759
## factor(Stress.Level)8                      -1.657444   3.165466  -0.524
## factor(Stress.Level)9                      -0.269294   2.631239  -0.102
## factor(Sleep.Quality)Fair:Duration..weeks. -0.051620   0.307122  -0.168
## factor(Sleep.Quality)Good:Duration..weeks. -0.001166   0.329358  -0.004
## factor(Sleep.Quality)Poor:Duration..weeks. -1.092225   0.317033  -3.445
## Duration..weeks.:factor(Stress.Level)2     -0.056465   0.326628  -0.173
## Duration..weeks.:factor(Stress.Level)3      0.222208   0.324192   0.685
## Duration..weeks.:factor(Stress.Level)4      0.058787   0.396866   0.148
## Duration..weeks.:factor(Stress.Level)5      0.568176   0.397329   1.430
## Duration..weeks.:factor(Stress.Level)6      0.042906   0.318948   0.135
## Duration..weeks.:factor(Stress.Level)7      0.312677   0.395000   0.792
```

```
## Duration..weeks.:factor(Stress.Level)8     -1.209145   0.398373  -3.035
## Duration..weeks.:factor(Stress.Level)9     -1.363695   0.326722  -4.174
##                                            Pr(>|t|)
## (Intercept)                                0.728340
## factor(Sleep.Quality)Fair                  0.977191
## factor(Sleep.Quality)Good                  0.763861
## factor(Sleep.Quality)Poor                  0.663273
## Duration..weeks.                           0.443521
## factor(Stress.Level)2                      0.983944
## factor(Stress.Level)3                      0.560692
## factor(Stress.Level)4                      0.864669
## factor(Stress.Level)5                      0.343901
## factor(Stress.Level)6                      0.698303
## factor(Stress.Level)7                      0.449961
## factor(Stress.Level)8                      0.602079
## factor(Stress.Level)9                      0.918752
## factor(Sleep.Quality)Fair:Duration..weeks. 0.866971
## factor(Sleep.Quality)Good:Duration..weeks. 0.997185
## factor(Sleep.Quality)Poor:Duration..weeks. 0.000932 ***
## Duration..weeks.:factor(Stress.Level)2     0.863210
## Duration..weeks.:factor(Stress.Level)3     0.495164
## Duration..weeks.:factor(Stress.Level)4     0.882633
## Duration..weeks.:factor(Stress.Level)5     0.156819
## Duration..weeks.:factor(Stress.Level)6     0.893345
## Duration..weeks.:factor(Stress.Level)7     0.431065
## Duration..weeks.:factor(Stress.Level)8     0.003290 **
## Duration..weeks.:factor(Stress.Level)9     7.89e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.822 on 76 degrees of freedom
## Multiple R-squared:  0.8897, Adjusted R-squared:  0.8563
## F-statistic: 26.65 on 23 and 76 DF,  p-value: < 2.2e-16
```

Our final model, before testing for the rest of our assumptions, will include the following predictors and interaction terms: Sleep.Quality, Duration..weeks., Stress.Level, Sleep.Quality:Duration..weeks., and Duration..weeks.:Stress.Level.

Based on this model, we can now test the rest of our assumptions. We will start with testing for linearity by inspecting our residual plots.

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

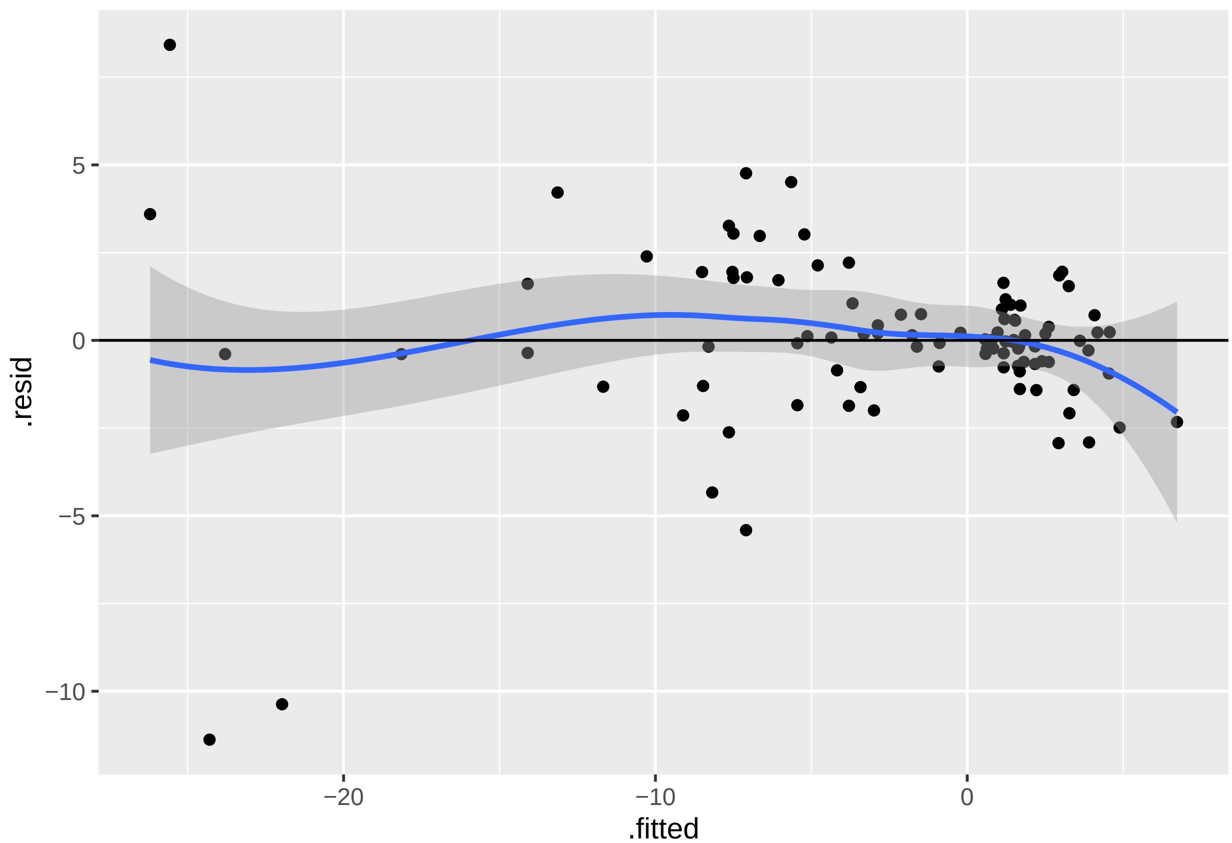


Based on our residuals, there appears to be a pattern, with most data points on the right side, as well as the presence of some funnel shape. We can try to correct this by adding some polynomial terms, or transforming the data. We will start with polynomial terms.

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ factor(Sleep.Quality) + Duration..weeks. +
##     I(Duration..weeks.^2) + factor(Stress.Level) + factor(Sleep.Quality) *
##     Duration..weeks. + Duration..weeks. * factor(Stress.Level),
##     data = weight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.3791  -0.7311  -0.0065   0.9976   8.4209
##
## Coefficients:
##                                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                                 1.66345    2.54941   0.652 0.516083
## factor(Sleep.Quality)Fair                   0.13760    2.13220   0.065 0.948715
## factor(Sleep.Quality)Good                  -0.74349    2.42611  -0.306 0.760108
## factor(Sleep.Quality)Poor                  -0.87654    2.00406  -0.437 0.663091
## Duration..weeks.                           -0.09704    0.48997  -0.198 0.843543
## I(Duration..weeks.^2)                       0.02828    0.02704   1.046 0.298892
## factor(Stress.Level)2                       0.11348    2.63001   0.043 0.965698
## factor(Stress.Level)3                      -1.46360    2.58117  -0.567 0.572388
## factor(Stress.Level)4                      -0.59995    3.09446  -0.194 0.846794
## factor(Stress.Level)5                      -2.49747    2.85091  -0.876 0.383813
```

```
## factor(Stress.Level)6                       -1.06770    2.49080  -0.429 0.669402
## factor(Stress.Level)7                       -2.34457    3.59394  -0.652 0.516159
## factor(Stress.Level)8                       -1.37165    3.17528  -0.432 0.666996
## factor(Stress.Level)9                       -0.38409    2.63190  -0.146 0.884363
## factor(Sleep.Quality)Fair:Duration..weeks. -0.09137    0.30928  -0.295 0.768475
## factor(Sleep.Quality)Good:Duration..weeks. -0.01614    0.32946  -0.049 0.961055
## factor(Sleep.Quality)Poor:Duration..weeks. -1.09414    0.31684  -3.453 0.000914
## Duration..weeks.:factor(Stress.Level)2     -0.06645    0.32656  -0.203 0.839304
## Duration..weeks.:factor(Stress.Level)3      0.22627    0.32401   0.698 0.487124
## Duration..weeks.:factor(Stress.Level)4      0.08923    0.39769   0.224 0.823079
## Duration..weeks.:factor(Stress.Level)5      0.55714    0.39722   1.403 0.164866
## Duration..weeks.:factor(Stress.Level)6      0.05918    0.31913   0.185 0.853376
## Duration..weeks.:factor(Stress.Level)7      0.27307    0.39657   0.689 0.493213
## Duration..weeks.:factor(Stress.Level)8     -1.23041    0.39864  -3.086 0.002837
## Duration..weeks.:factor(Stress.Level)9     -1.36556    0.32652  -4.182 7.75e-05
##
## (Intercept)
## factor(Sleep.Quality)Fair
## factor(Sleep.Quality)Good
## factor(Sleep.Quality)Poor
## Duration..weeks.
## I(Duration..weeks.^2)
## factor(Stress.Level)2
## factor(Stress.Level)3
## factor(Stress.Level)4
## factor(Stress.Level)5
## factor(Stress.Level)6
## factor(Stress.Level)7
## factor(Stress.Level)8
## factor(Stress.Level)9
## factor(Sleep.Quality)Fair:Duration..weeks.
## factor(Sleep.Quality)Good:Duration..weeks.
## factor(Sleep.Quality)Poor:Duration..weeks. ***
## Duration..weeks.:factor(Stress.Level)2
## Duration..weeks.:factor(Stress.Level)3
## Duration..weeks.:factor(Stress.Level)4
## Duration..weeks.:factor(Stress.Level)5
## Duration..weeks.:factor(Stress.Level)6
## Duration..weeks.:factor(Stress.Level)7
## Duration..weeks.:factor(Stress.Level)8     **
## Duration..weeks.:factor(Stress.Level)9     ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.82 on 75 degrees of freedom
## Multiple R-squared:  0.8913, Adjusted R-squared:  0.8565
## F-statistic: 25.62 on 24 and 75 DF,  p-value: < 2.2e-16

## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



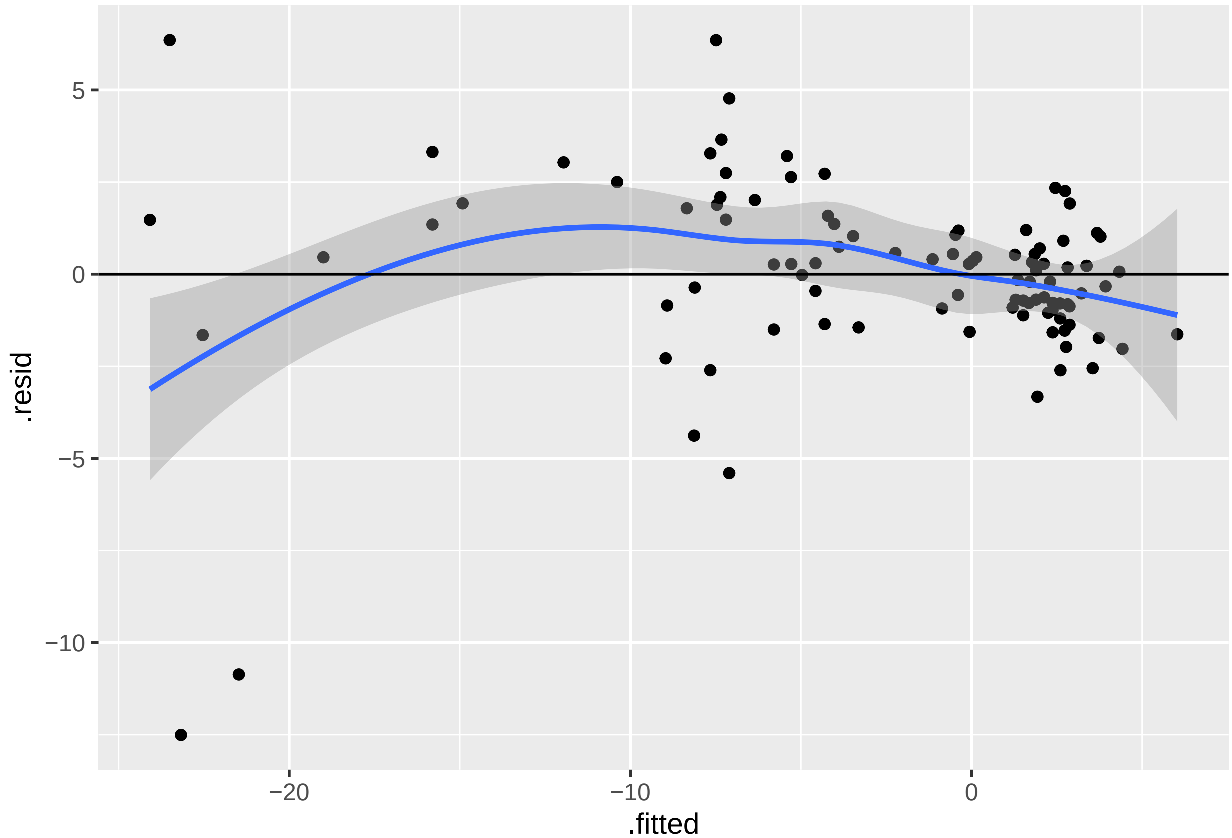
Adding a polynomial term appears to not be the solution, as not only does the residual plot appear the same, but the adjust R-sqaured barely increases. There is no point to preventing our interpretation for such a small increase.

Next we will try performing a log transformation on Duration..weeks.

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ factor(Sleep.Quality) + log(Duration..weeks.) +
##     factor(Stress.Level) + factor(Sleep.Quality) * log(Duration..weeks.) +
##     log(Duration..weeks.) * factor(Stress.Level), data = weight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.5056  -0.9419   0.2044   1.2345   6.3533
##
## Coefficients:
##                                              Estimate Std. Error t value
## (Intercept)                                   0.27355    3.42430   0.080
## factor(Sleep.Quality)Fair                     0.11340    2.82176   0.040
## factor(Sleep.Quality)Good                    -2.57483    3.44092  -0.748
## factor(Sleep.Quality)Poor                     1.58927    2.86099   0.555
## log(Duration..weeks.)                         1.28194    1.93076   0.664
## factor(Stress.Level)2                         0.81976    3.55617   0.231
## factor(Stress.Level)3                        -2.72618    3.42733  -0.795
## factor(Stress.Level)4                        -2.26130    3.85645  -0.586
## factor(Stress.Level)5                        -4.49656    4.26194  -1.055
```

```
## factor(Stress.Level)6                          -0.34971    3.35196  -0.104
## factor(Stress.Level)7                          -2.58341    5.06097  -0.510
## factor(Stress.Level)8                           2.24593    4.16139   0.540
## factor(Stress.Level)9                           0.06974    3.32780   0.021
## factor(Sleep.Quality)Fair:log(Duration..weeks.) -0.08017    1.58629  -0.051
## factor(Sleep.Quality)Good:log(Duration..weeks.)  0.94134    1.84057   0.511
## factor(Sleep.Quality)Poor:log(Duration..weeks.) -5.61445    1.68822  -3.326
## log(Duration..weeks.):factor(Stress.Level)2    -0.63803    1.84765  -0.345
## log(Duration..weeks.):factor(Stress.Level)3     1.58105    1.79470   0.881
## log(Duration..weeks.):factor(Stress.Level)4     1.03392    2.02085   0.512
## log(Duration..weeks.):factor(Stress.Level)5     3.12719    2.36324   1.323
## log(Duration..weeks.):factor(Stress.Level)6    -0.05704    1.79414  -0.032
## log(Duration..weeks.):factor(Stress.Level)7     1.26398    2.41255   0.524
## log(Duration..weeks.):factor(Stress.Level)8    -6.77947    2.11771  -3.201
## log(Duration..weeks.):factor(Stress.Level)9    -6.13711    1.75701  -3.493
##                                                Pr(>|t|)
## (Intercept)                                     0.93654
## factor(Sleep.Quality)Fair                       0.96805
## factor(Sleep.Quality)Good                       0.45659
## factor(Sleep.Quality)Poor                       0.58019
## log(Duration..weeks.)                           0.50873
## factor(Stress.Level)2                           0.81831
## factor(Stress.Level)3                           0.42884
## factor(Stress.Level)4                           0.55936
## factor(Stress.Level)5                           0.29475
## factor(Stress.Level)6                           0.91718
## factor(Stress.Level)7                           0.61121
## factor(Stress.Level)8                           0.59098
## factor(Stress.Level)9                           0.98333
## factor(Sleep.Quality)Fair:log(Duration..weeks.)  0.95983
## factor(Sleep.Quality)Good:log(Duration..weeks.)  0.61053
## factor(Sleep.Quality)Poor:log(Duration..weeks.)  0.00136 **
## log(Duration..weeks.):factor(Stress.Level)2     0.73081
## log(Duration..weeks.):factor(Stress.Level)3     0.38112
## log(Duration..weeks.):factor(Stress.Level)4     0.61040
## log(Duration..weeks.):factor(Stress.Level)5     0.18971
## log(Duration..weeks.):factor(Stress.Level)6     0.97472
## log(Duration..weeks.):factor(Stress.Level)7     0.60186
## log(Duration..weeks.):factor(Stress.Level)8     0.00200 **
## log(Duration..weeks.):factor(Stress.Level)9     0.00080 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.917 on 76 degrees of freedom
## Multiple R-squared:  0.8821, Adjusted R-squared:  0.8464
## F-statistic: 24.72 on 23 and 76 DF,  p-value: < 2.2e-16
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



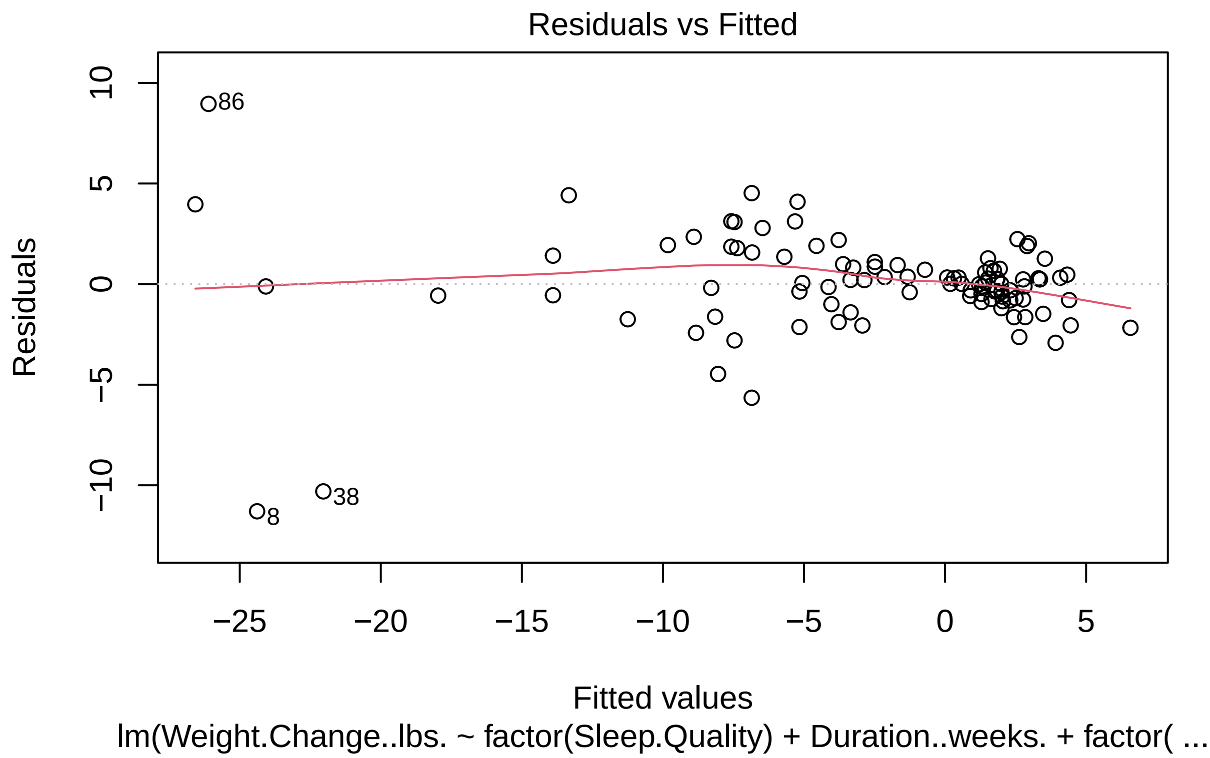
Transforming Duration..weeks. using a log transformation appears to not correct the linearity problem either. In fact the residual plot now looks worse than before.

Based on these attempts, we may need to add more complex polynomial or transformations to our data. We may also need to attempt some other model types, but for the sake of this class we will move forward using our model before we attempting correcting linearity, and simply state that our model fails the linearity assumption.

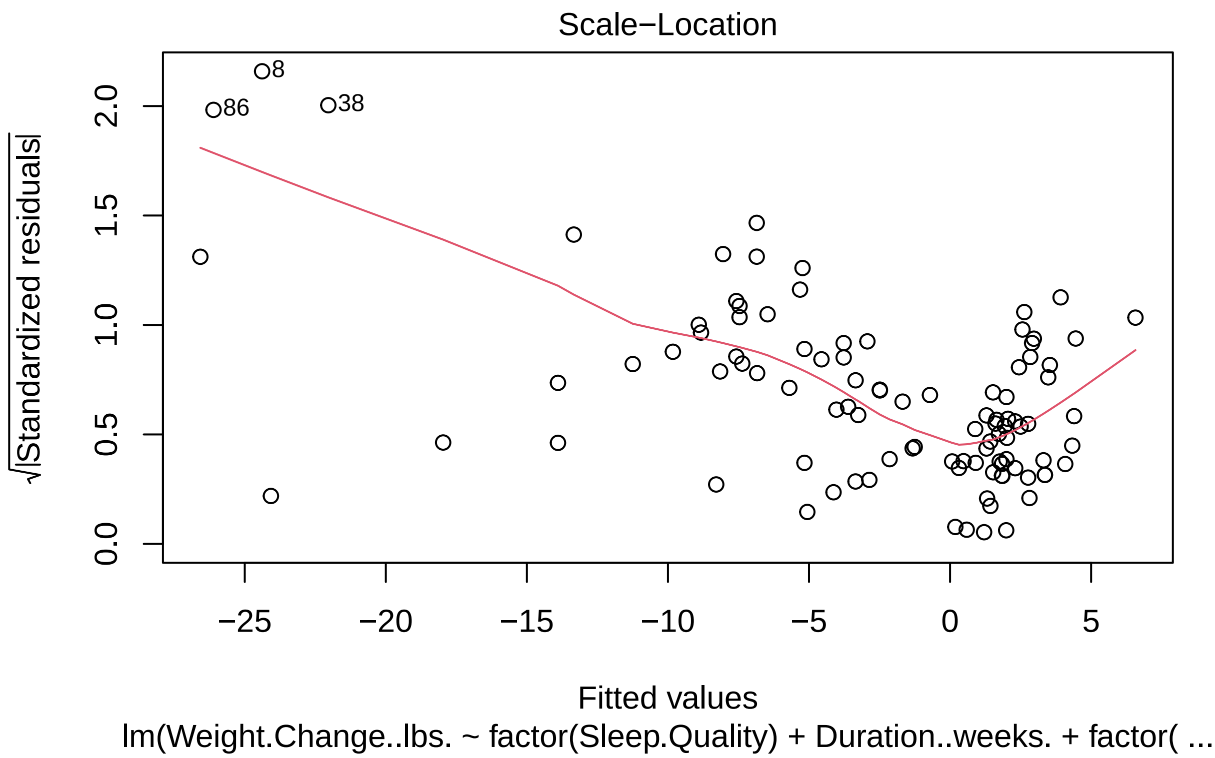
The next assumption we will look at is independence. As our responding variable, Weight.Change..lbs. is not considered time-series data we can conclude this assumption has been passed.

To test for homoscedasticity we can look at our residual plots, scale-location plots, and run a Breusch-Pagan Test, using the following hypothesis:

$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_n^2$ (heteroscedasticity is not present)H_a : at least one σ_i^2 is different from the others i = 1, 2, ..., n (het

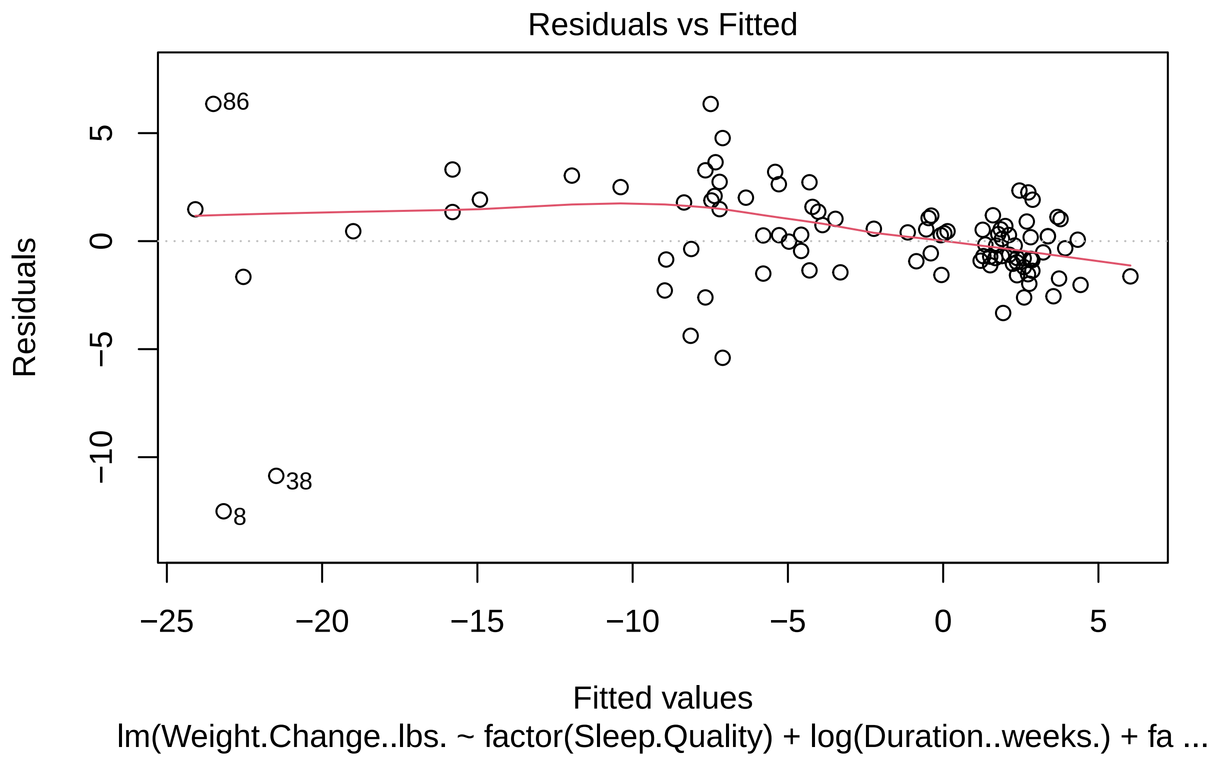


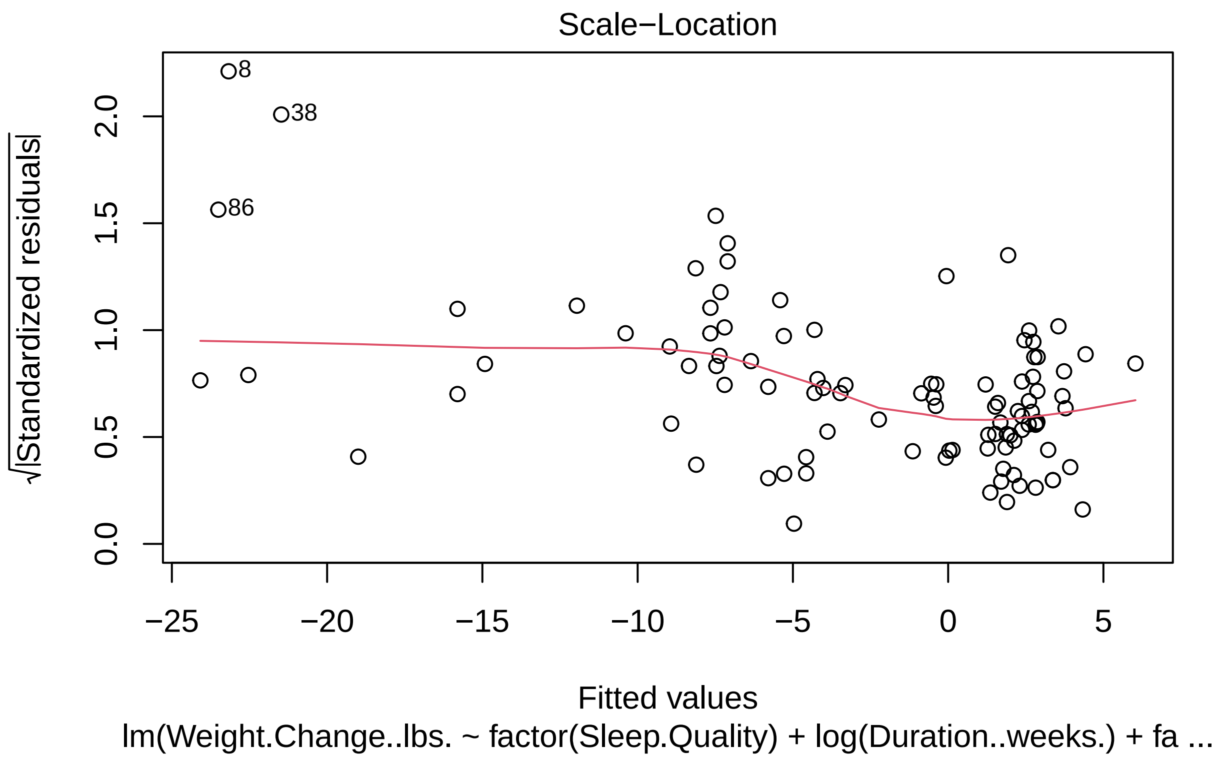
Residuals vs Fitted
Residuals
Fitted values
lm(Weight.Change..lbs. ~ factor(Sleep.Quality) + Duration..weeks. + factor(...
86
8
38
10
5
0
−5
−10
−25
−20
−15
−10
−5
0
5




```
##
##  studentized Breusch-Pagan test
##
## data:  weight_model_int_final
## BP = 44.674, df = 23, p-value = 0.004356
```

Based on our plots and the fact that our p-value for our Breusch-Pagan Test (0.004356) is < 0.05, we can reject our null hypothesis, implying the presence of heteroscedasticity. To try and correct this, we can attempt a log transformation.


Residuals vs Fitted
86
38
8
5
0
−5
−10
Residuals
−25
−20
−15
−10
−5
0
5
Fitted values
lm(Weight.Change..lbs. ~ factor(Sleep.Quality) + log(Duration..weeks.) + fa ...

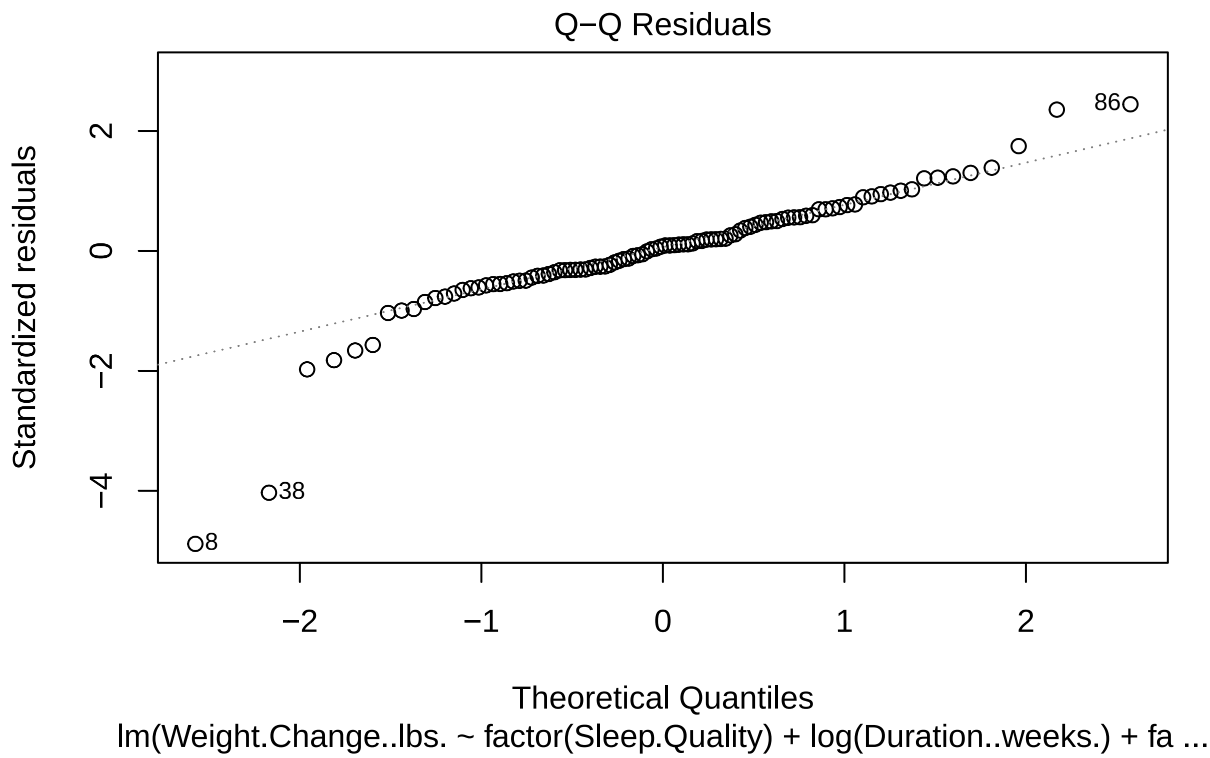



```
##
##  studentized Breusch-Pagan test
##
## data:  weight_model_log
## BP = 30.059, df = 23, p-value = 0.1477
```

Our plots look much better now than before, and we can see that our p-value for our Breusch-Pagan Test (0.1477) is > 0.05. This means that we now fail to reject our null hypotheis, implying heteroscedasticity is not present. This suggests that our log transformation has improved our model, allowing it to now pass the homoscedasticity assumption. It should be noted that the adjusted R-square of our log model has decreased slightly from our previous model (0.8464 down from 0.8563). Despite this loss, we are comfortable moving forward with the log model, as it helps us pass one of our assumptions that was previously not met.

To test whether our residuals are normally distributed, we can inspect our Q-Q plot, as well as run a Shapiro-Wilk test, using the following hypothesis:

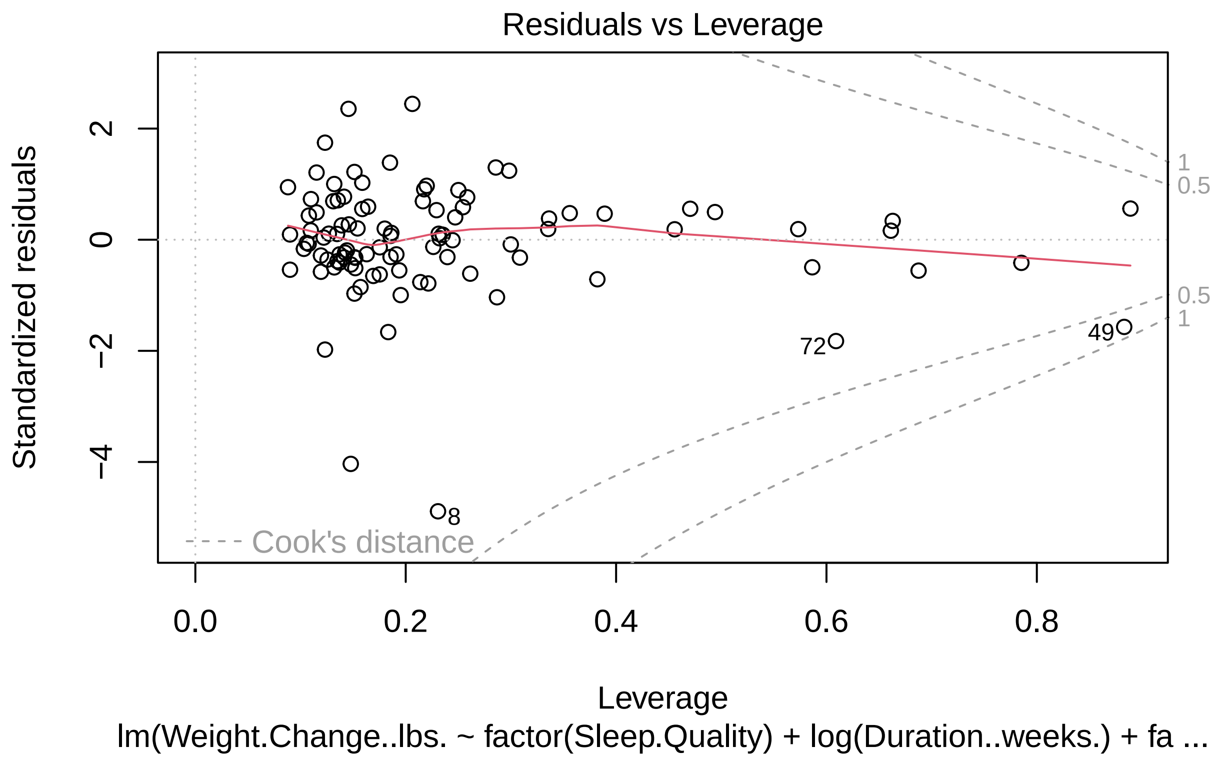
$$H_0 : \text{the residuals are normally distributed} H_a : \text{the residuals are not normally distributed}$$




```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(weight_model_log)
## W = 0.85478, p-value = 1.788e-08
```

Based on our Q-Q plot, we can see that the residuals are likely not normally distributed. This is further reinforced by our Shapiro Wilks test, which returned a p-value 1.788e-08. As this is < 0.05, we reject our null hypothesis, implying that the residuals are not normally distributed. We can attempt to correct this using a Box-Cox transformation. Unfortunately, as our responding variable is weight change, there are values less than 0. Therefore a Box-Cox transformation cannot be done. For the time being we will simply conclude that our model has not passed the assumption of normality.

Finally, we will check for outliers, using our residuals vs. leverage plot.



Based on our plot, we can see that observation 49 is an outlier. We can attempt to remove the data point to see how it affects our model.

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ factor(Sleep.Quality) + log(Duration..weeks.) +
##     factor(Stress.Level) + factor(Sleep.Quality) * log(Duration..weeks.) +
##     log(Duration..weeks.) * factor(Stress.Level), data = weight_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.7111  -0.8341   0.1531   1.1537   8.3484
##
## Coefficients:
##                                            Estimate Std. Error t value
## (Intercept)                                 0.21766    3.39093   0.064
## factor(Sleep.Quality)Fair                   0.18485    2.79447   0.066
## factor(Sleep.Quality)Good                  -1.59689    3.46262  -0.461
## factor(Sleep.Quality)Poor                   1.24320    2.84135   0.438
## log(Duration..weeks.)                       1.21315    1.91233   0.634
## factor(Stress.Level)2                       0.54944    3.52544   0.156
## factor(Stress.Level)3                      -2.44803    3.39827  -0.720
## factor(Stress.Level)4                      -1.93746    3.82412  -0.507
## factor(Stress.Level)5                      -4.19952    4.22433  -0.994
## factor(Stress.Level)6                      -0.32160    3.31916  -0.097
## factor(Stress.Level)7                      -2.49840    5.01166  -0.499
## factor(Stress.Level)8                      13.14533    8.01703   1.640
```

```
## factor(Stress.Level)9                          -0.12103    3.29738  -0.037
## factor(Sleep.Quality)Fair:log(Duration..weeks.)  -0.06075    1.57079  -0.039
## factor(Sleep.Quality)Good:log(Duration..weeks.)   0.55598    1.83868   0.302
## factor(Sleep.Quality)Poor:log(Duration..weeks.)  -5.29987    1.68341  -3.148
## log(Duration..weeks.):factor(Stress.Level)2     -0.46614    1.83275  -0.254
## log(Duration..weeks.):factor(Stress.Level)3      1.45515    1.77888   0.818
## log(Duration..weeks.):factor(Stress.Level)4      0.89513    2.00296   0.447
## log(Duration..weeks.):factor(Stress.Level)5      2.98467    2.34180   1.275
## log(Duration..weeks.):factor(Stress.Level)6     -0.02623    1.77666  -0.015
## log(Duration..weeks.):factor(Stress.Level)7      1.23485    2.38898   0.517
## log(Duration..weeks.):factor(Stress.Level)8    -12.05261    3.93280  -3.065
## log(Duration..weeks.):factor(Stress.Level)9     -6.05003    1.74066  -3.476
##                                                Pr(>|t|)
## (Intercept)                                    0.948991
## factor(Sleep.Quality)Fair                      0.947436
## factor(Sleep.Quality)Good                      0.646004
## factor(Sleep.Quality)Poor                      0.662978
## log(Duration..weeks.)                          0.527760
## factor(Stress.Level)2                          0.876570
## factor(Stress.Level)3                          0.473533
## factor(Stress.Level)4                          0.613893
## factor(Stress.Level)5                          0.323358
## factor(Stress.Level)6                          0.923071
## factor(Stress.Level)7                          0.619579
## factor(Stress.Level)8                          0.105261
## factor(Stress.Level)9                          0.970818
## factor(Sleep.Quality)Fair:log(Duration..weeks.) 0.969254
## factor(Sleep.Quality)Good:log(Duration..weeks.) 0.763198
## factor(Sleep.Quality)Poor:log(Duration..weeks.) 0.002357 **
## log(Duration..weeks.):factor(Stress.Level)2    0.799930
## log(Duration..weeks.):factor(Stress.Level)3    0.415939
## log(Duration..weeks.):factor(Stress.Level)4    0.656233
## log(Duration..weeks.):factor(Stress.Level)5    0.206415
## log(Duration..weeks.):factor(Stress.Level)6    0.988259
## log(Duration..weeks.):factor(Stress.Level)7    0.606752
## log(Duration..weeks.):factor(Stress.Level)8    0.003028 **
## log(Duration..weeks.):factor(Stress.Level)9    0.000851 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 2.889 on 75 degrees of freedom
## Multiple R-squared:  0.8859, Adjusted R-squared:  0.8509
## F-statistic: 25.32 on 23 and 75 DF,  p-value: < 2.2e-16
```

Based on the output of our model with and without the outlier, we can see that removing the outlier slightly improved our adjusted R-squared (0.8509 from 0.8464), while reducing or RSE (2.889 from 2.917), without changing the significance of any of our predictors. Based on this we feel comfortable removing the indicated outlier.

Although our model does fail the assumptions of linearity and normality, for the sake of this project we will go forward with our final model being:

$$X_{Weight.Change..lbs.} = 0.21766 + 0.18485 X_{Sleep.Quality_{Fair}} - 1.59689 X_{Sleep.Quality_{Good}} + 1.24320 X_{Sleep.Quality_{Poor}} + 1.21315 X_{log(}$$

Now that we have our final model, we can begin interpreting it.

Intercept: When sleep quality is excellent, stress level is 1, and log(duration) is 0 (which would be a duration of 1 week), the expected weight change is 0.21766 pounds.

Sleep Quality:

- If sleep quality is fair, the predicted weight change is $(0.18485 - 0.06075X_{log(Duration..weeks.)})$ pounds.
- If sleep quality is good, the predicted weight change is $(-1.59689 + 0.55598X_{log(Duration..weeks.)})$ pounds.
- If sleep quality is poor, the predicted weight change is $(1.24320 - 5.29987X_{log(Duration..weeks.)})$ pounds.
- If sleep quality is excellent, the predicted weight change would simply be our model equation excluding any $Sleep.Quality_i$ terms or interactions, where i = Fair, Good, or Poor, as all values would simply be 0 for those variables.

Duration in weeks:

- For every one-unit increase in log(Duration..weeks.), weight change increases by:

  $(1.21315 - 0.06075X_{Sleep.Quality_{Fair}} + 0.55598X_{Sleep.Quality_{Good}} - 5.29987X_{Sleep.Quality_{Poor}} - 0.46614X_{Stress.Level_2} + 1.4551$

Stress Level:

- If stress level is 2, the predicted weight change is $(0.54944 - 0.46614X_{log(Duration..weeks.)})$ pounds.
- If stress level is 3, the predicted weight change is $(-2.44803 + 1.45515X_{log(Duration..weeks.)})$ pounds.
- If stress level is 4, the predicted weight change is $(-1.93746 + 0.89513X_{log(Duration..weeks.)})$ pounds.
- If stress level is 5, the predicted weight change is $(-4.19952 + 2.98467X_{log(Duration..weeks.)})$ pounds.
- If stress level is 6, the predicted weight change is $(-0.32160 - 0.02623X_{log(Duration..weeks.)})$ pounds.
- If stress level is 7, the predicted weight change is $(-2.49840 + 1.23485X_{log(Duration..weeks.)})$ pounds.
- If stress level is 8, the predicted weight change is $(13.14533 - 12.05261X_{log(Duration..weeks.)})$ pounds.
- If stress level is 9, the predicted weight change is $(-0.12103 - 6.05003X_{log(Duration..weeks.)})$ pounds.
- If stress level is 1, the predicted weight change would simply be our model equation excluding any $Stress.Level_i$ terms or interactions, where i = 2 to 9, as all values would simply be 0 for those variables.

Lastly, our Adjusted R^2 for our model is 0.8509, implying that our 85.09% of the variability in weight change is explained by our model.

To demonstrate the predictiveness of our model, we have included an example. Assume a participant records their weight change over a period of 6 weeks, where they recorded poor sleep quality, and a stress level of 5. What would be there predicted weight change?

```
##         1
## -4.713261
```

Based on our finalized model, and assuming a participant records their weight change over a period of 6 weeks, where they recorded poor sleep quality, and a stress level of 5, we would predict their weight to decrease by 4.713261 pounds.

4 Conclusion and Discussion

4.1 Approach

4.2 Future Work

5 References

1. THE WEBSITE FOR THE DATASET