The Relationship Between Exercise Time and Weight Loss

Bilal Mozaffar, Risheng Li, Thomas Janes

2/6/2021

```
## Warning: package 'tidyr' was built under R version 3.6.2

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

##
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':
##
##     smiths
```

Abstract

The purpose of this data analysis is to apply statistical models to health data for a call center. While there were many data points missing for weight gain and the total metabolic minutes, many of these could be found using the other metrics available. The health data was processed to provide as many viable data points as possible without compromising the integrity of the analysis. Two models were constructed: a linear model to examine the relationship between total metabolic minutes and weight gain and a logistic regression model to examine the relationship between shift time and weight gain.

Introduction

We have been given several health measurements of employees working at a call center. Over eight months, the metrics have tracked information including work shift time, exercise time, weight gained (binary yes/no measurement), amount of weight gained, total metabolic minutes, and more. We hope to figure out whether certain factors, specifically total metabolic minutes and shift time, play a role in employees' weight gain.

The Data

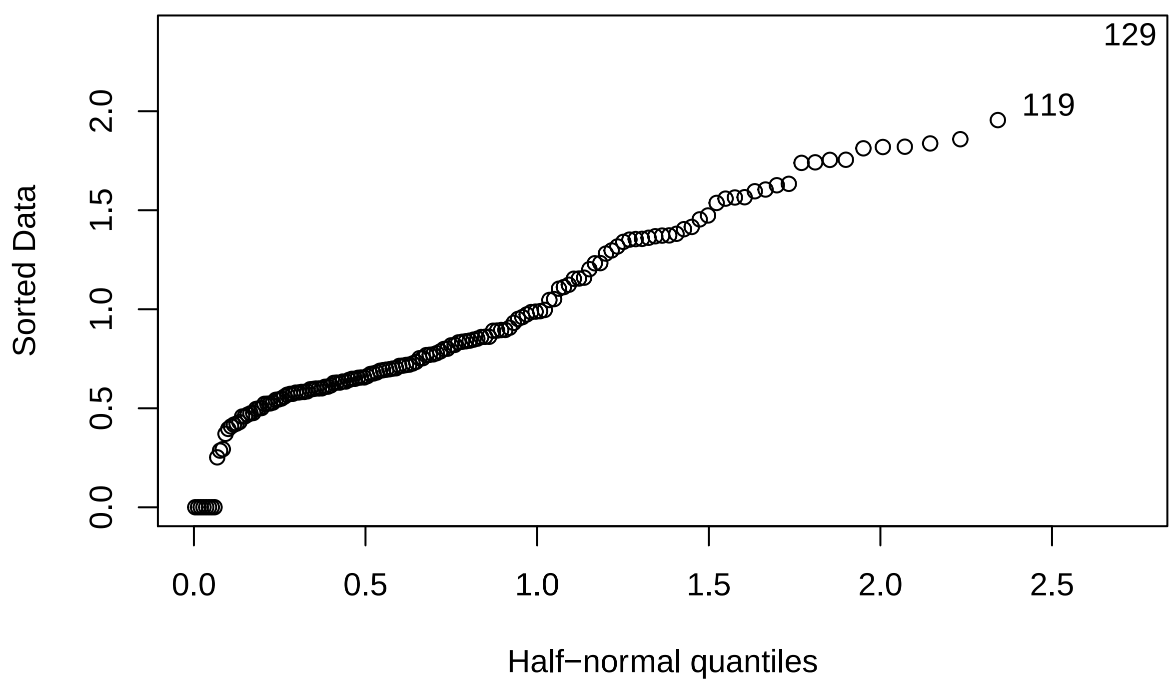
The data used in this analysis was provided by the call center. The data consists of 392 observations and 83 features, with variables such as member information, body measurements, exercise time, and weight situation. There are thirteen variables that we will use for this analysis, including shift time, gender, age, height, weight gain, pounds gained, body weight, body mass index, vigorous exercise time, moderate exercise time, walking exercise time, and total metabolic minutes. It is important to note that there were many missing values for total metabolic minutes, but these could be filled in using the other variables available. There were also missing values for pounds gained, but no features were available to account for this so these observations were removed.

We load and subset the raw data into a new dataset that contains the columns that we will use for our analysis. After the initial examination of the data, we renamed several variables by eliminating white spaces in order to improve the easiness of analysis. Using the formula provided, Total_met_min = 8 * Vig_ex_time + 4 * Mod_ex_time + 3.3 * Walk_ex_time, we filled in the missing values of variable Total_Met_Min.

replace lbs_gained value with 0 if weightgain = 0 and drop the rows that both weightgain and lbs_gained are missing. Also calculate the beginning weight by subtracting lbs_gain from body weight. Calculate the change in BMI using begeinning weight and height.

By turning weightgain into a dummy variable and running a logistic regression, we can measure if other variables have influence on weightgain. First, we factored the two categorical variables and built a model with variables include shift, gender, Age, height and Beg_BMI and Total_Met_Min. Because the change in BMI is very minimal, we decided to use the beginning BMI as one of the regressors. The half normal plot shows that there is no obvious outliers in the model.

```
## Warning: package 'faraway' was built under R version 3.6.3
```



Solely from the model, we can see that only three variables have significant impact on the predictor: gender, Beg_BMI and Total_Met_Min. We perform a Chi-sqaure test to test if the model contains shift is better than the model without shifts. The null hypothesis is that the model without shift is a better model, and the alternative hypothesis is that the model without shift is not a better model. In order to compare two models, we omit the NA values in the dataset to make sure the number of cases used in each model is the same. The test statistic is is very large given a 95% confidence interval, therefore we fail to reject the null hypothesis that the reduced model is better, which means that the model contains shift is less favorable than the reduced model.

```
##
## Call:
## glm(formula = weightgain.b ~ shift + gender + Age + height +
##     Beg_BMI + Total_Met_Min, family = binomial, data = na.omit(dat2))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3879  -0.2230   0.5824   0.7568   1.4539
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)    3.116e+01  4.840e+03   0.006    0.995
## shift10am     -1.711e+01  2.788e+03  -0.006    0.995
## shift11am     -1.759e+01  2.788e+03  -0.006    0.995
## shift12pm     -1.721e+01  2.788e+03  -0.006    0.995
## shift1pm      -1.735e+01  2.788e+03  -0.006    0.995
## shift2pm      -1.544e+00  3.271e+03   0.000    1.000
## shift7am      -1.747e+01  2.788e+03  -0.006    0.995
## shift8am      -1.785e+01  2.788e+03  -0.006    0.995
## shift9am      -1.760e+01  2.788e+03  -0.006    0.995
## shiftother    -1.731e+01  2.788e+03  -0.006    0.995
## genderFemale  -1.569e+01  3.956e+03  -0.004    0.997
## genderMale    -1.689e+01  3.956e+03  -0.004    0.997
## Age            2.771e-02  2.183e-02   1.269    0.204
## height         6.195e-02  6.587e-02   0.940    0.347
## Beg_BMI       -7.916e-02  3.692e-02  -2.144    0.032 *
## Total_Met_Min  6.699e-03  4.172e-03   1.606    0.108
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 175.45  on 155  degrees of freedom
## Residual deviance: 155.29  on 140  degrees of freedom
## AIC: 187.29
##
## Number of Fisher Scoring iterations: 16
```

```
##
## Call:
## glm(formula = weightgain.b ~ gender + Age + height + Beg_BMI +
##     Total_Met_Min, family = binomial, data = na.omit(dat2))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

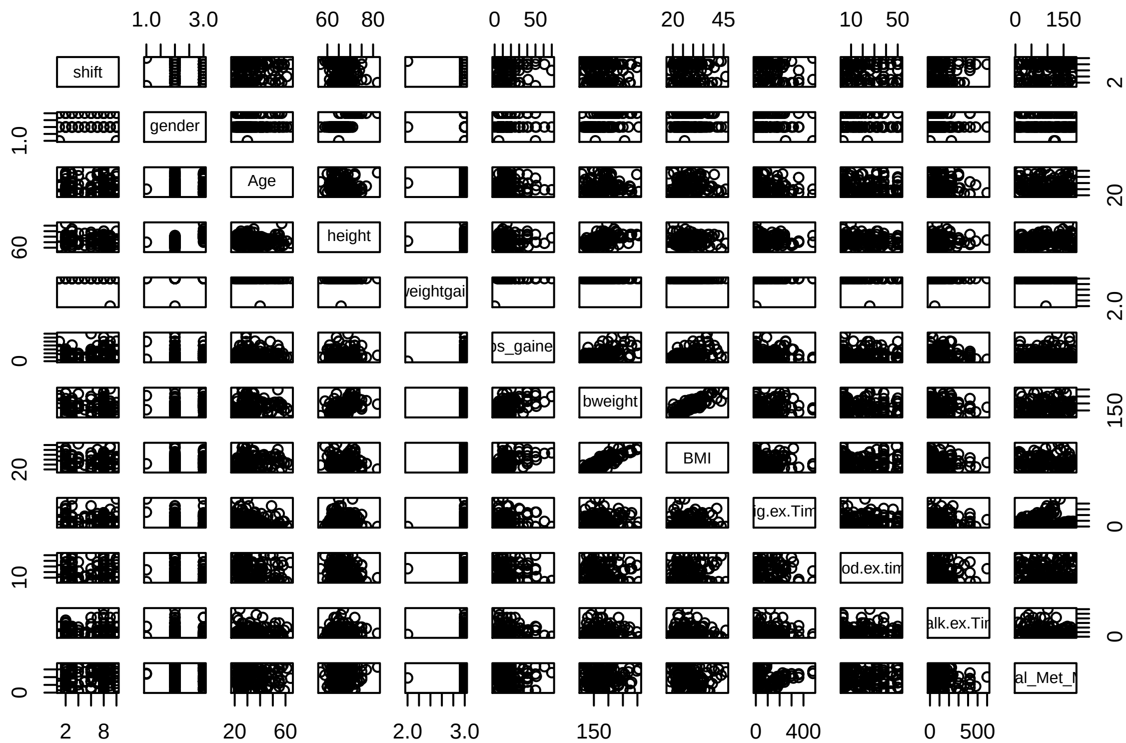
```
## -2.1953  -0.2586   0.5782   0.7459   1.3439
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.150e+01  1.455e+03   0.008    0.994
## genderFemale  -1.390e+01  1.455e+03  -0.010    0.992
## genderMale    -1.515e+01  1.455e+03  -0.010    0.992
## Age            2.501e-02  2.075e-02   1.205    0.228
## height         6.508e-02  6.404e-02   1.016    0.310
## Beg_BMI       -6.662e-02  3.429e-02  -1.943    0.052 .
## Total_Met_Min  6.042e-03  3.913e-03   1.544    0.123
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 175.45  on 155  degrees of freedom
## Residual deviance: 161.64  on 149  degrees of freedom
## AIC: 175.64
##
## Number of Fisher Scoring iterations: 14
```

```
## Analysis of Deviance Table
##
## Model 1: weightgain.b ~ gender + Age + height + Beg_BMI + Total_Met_Min
## Model 2: weightgain.b ~ shift + gender + Age + height + Beg_BMI + Total_Met_Min
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       149     161.65
## 2       140     155.29  9   6.3589   0.7035
```

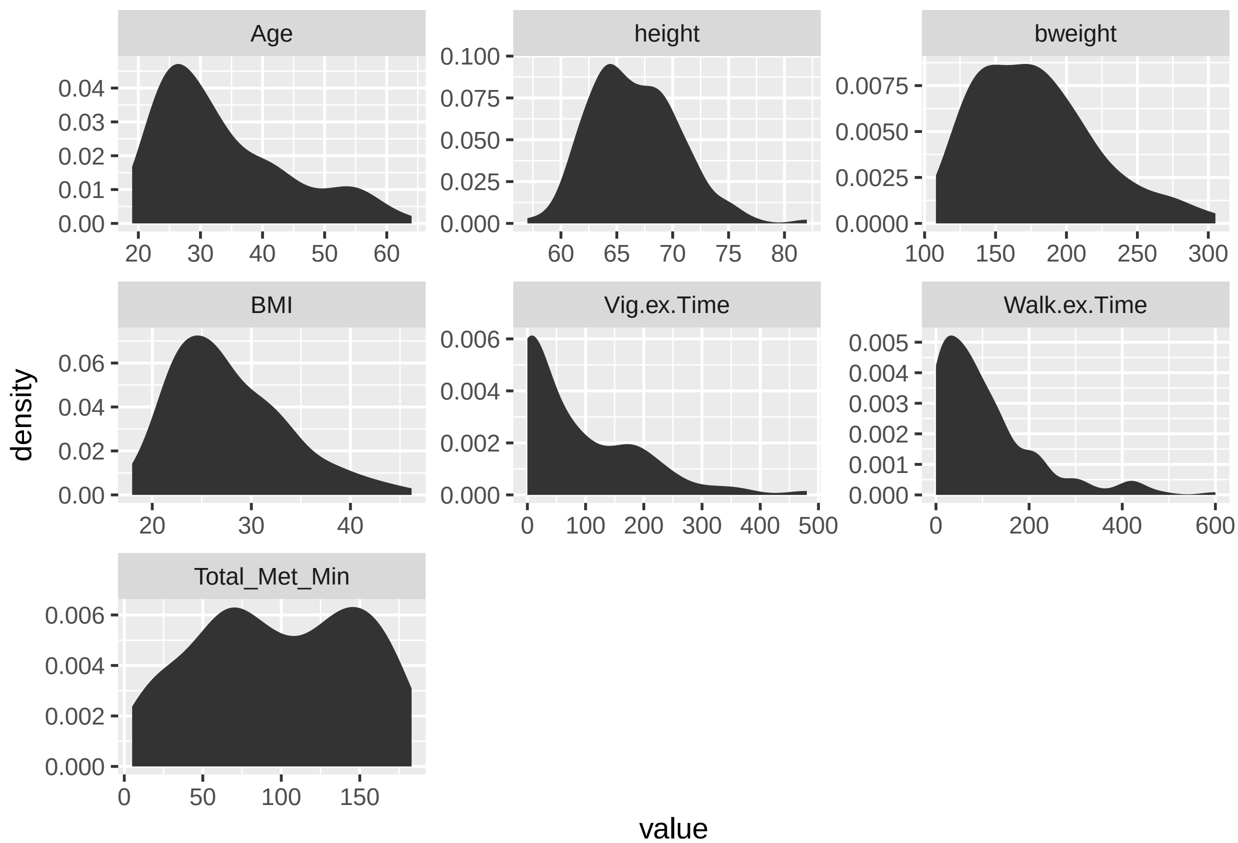
Since the initial model shows that Total_Met_Min is a significant variable, we performed another analysis, using a forward stepwise selection to select the model with the most appropriate variables that produces the lowest AIC. The selected variables are exactly the same as our previous analysis, which are gender, Beg_BMI and Total_Met_Min, which reassures that total metabolic minutes do have an effect on weight gain and shift does not have an effect on weightgain.

```
##              Step Df Deviance Resid. Df Resid. Dev      AIC
## 1                 NA       NA       155   175.4486 177.4486
## 2       + Beg_BMI -1 4.342492       154   171.1061 175.1061
## 3        + gender -2 4.834751       152   166.2713 174.2713
## 4 + Total_Met_Min -1 2.372421       151   163.8989 173.8989
```

```
##
## Call:  glm(formula = weightgain.b ~ Beg_BMI + gender + Total_Met_Min,
##     family = binomial, data = na.omit(dat2))
##
## Coefficients:
##   (Intercept)        Beg_BMI   genderFemale     genderMale  Total_Met_Min
##     15.439615      -0.064131     -12.835663     -13.706634       0.005863
##
## Degrees of Freedom: 155 Total (i.e. Null);  151 Residual
## Null Deviance:       175.4
## Residual Deviance: 163.9     AIC: 173.9
```



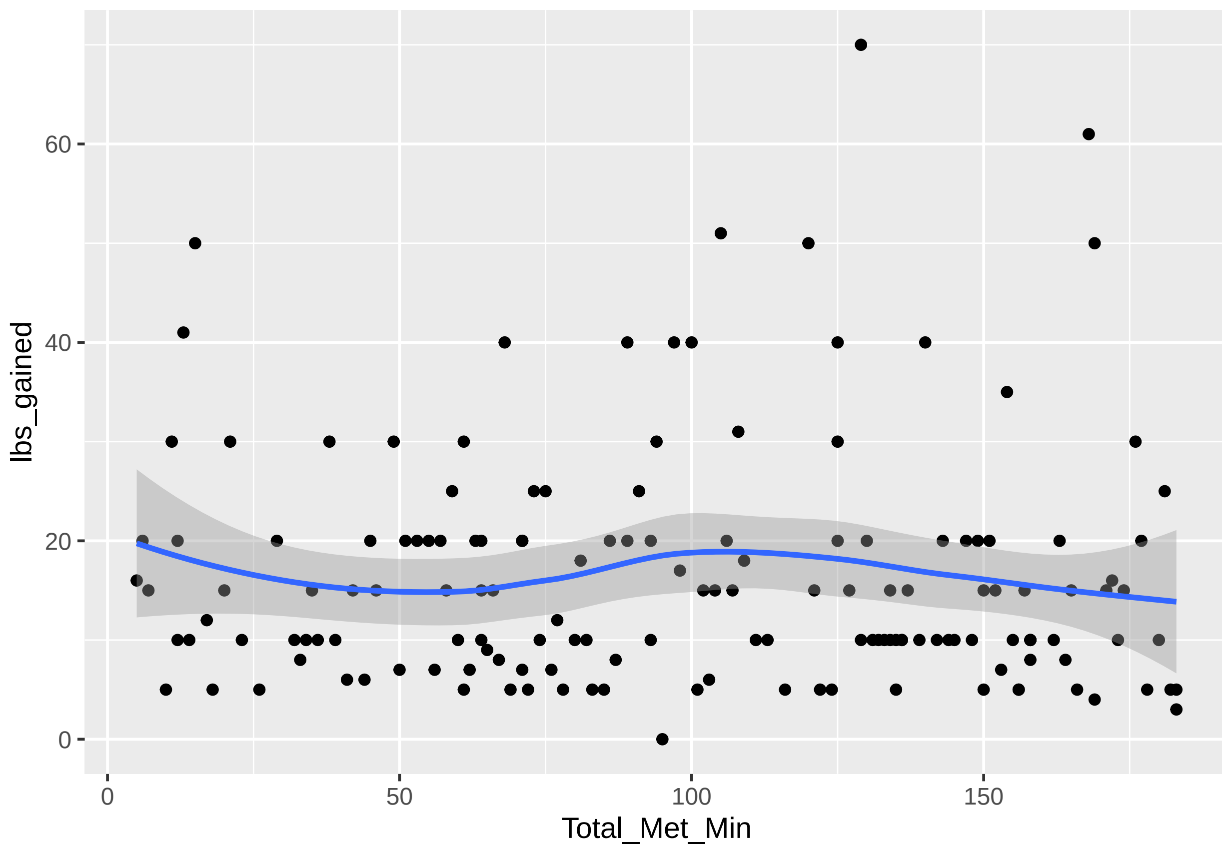

```
## Using Mod.ex.time as id variables

## Warning: Removed 82 rows containing non-finite values (stat_density).
```



We see that Age and BMI are right-skewed with nonzero values, so we will do a log transformation on these variables.

```
##
## Call:
## lm(formula = lbs_gained ~ Total_Met_Min, data = dat1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.406  -7.288  -1.913   3.543  53.824
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.048084   2.115141   8.060 2.03e-13 ***
## Total_Met_Min -0.006761   0.019050  -0.355    0.723
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.05 on 153 degrees of freedom
## Multiple R-squared:  0.0008225,  Adjusted R-squared:  -0.005708
## F-statistic: 0.126 on 1 and 153 DF,  p-value: 0.7232
```

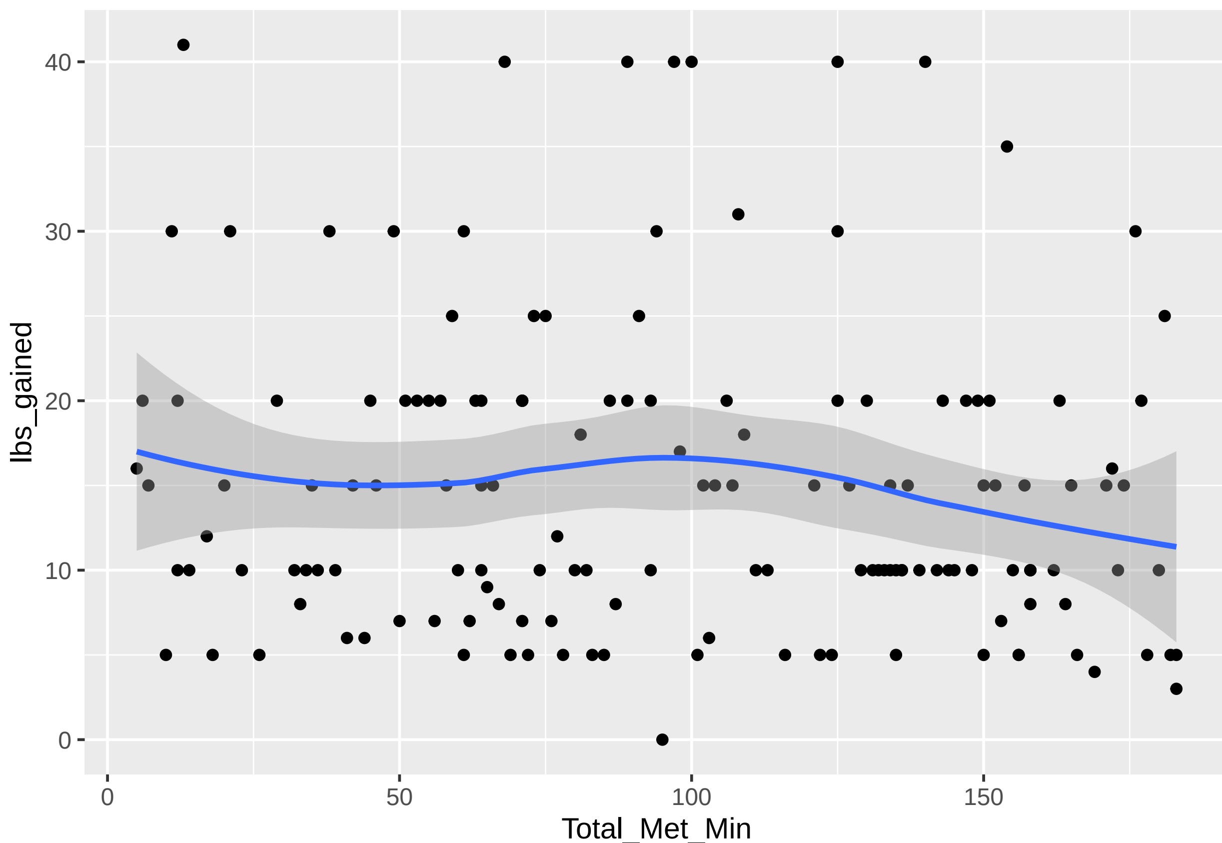
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
##
## Call:
## lm(formula = lbs_gained ~ Total_Met_Min, data = dat1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.875  -6.588  -3.225   4.470  26.077
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   16.88414    1.62851  10.368   <2e-16 ***
## Total_Met_Min -0.02115    0.01477  -1.432    0.154
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.127 on 147 degrees of freedom
## Multiple R-squared:  0.01376,    Adjusted R-squared:  0.007054
## F-statistic: 2.051 on 1 and 147 DF,  p-value: 0.1542
```

P-value got bigger

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



The Relationship Between Total Metabolic Minutes and Weight Gain

Using various methods of analysis, we found that total metabolic minutes is a significant variable. The initial model demonstrated that there is a negative relationship between total metabolic minutes and weight gain, and this was corroborated by the subsequent forward stepwise model.

The Relationship Between Shift Time and Weight Gain

It does not appear that shift has significant impact on weight gain. From the Chi-square test, we found that the model without shift performs better than the model with shift. As mentioned, we failed to reject the null hypothesis that the reduced model is better, so we do not find it beneficial to include shift in our final model to predict weight gain.

Conclusion

Our initial goal was to see whether shift and total metabolic minutes have an impact on weight gain. Multiple forms of analysis indicate that shift is not a relevant factor in determining weight gain, while total metabolic minutes is. Total metabolic minutes has a slightly adverse relationship to weight gain. We conducted further analysis to go beyond the requested covariate relationships. For a final model selection, gender, Beg_BMI, and Total_Met_Min appear to be the most signficant and useful predictors of weight gain.