The Effect of Total Metabolic Minutes and Start Shift Time on Binary and Numerical Weight Gain

Bilal Mozaffar, Risheng Li, Thomas Janes

2/6/2021

Abstract

The purpose of this data analysis is to determine whether and how total metabolic minutes and employee shift time affect weight gain. The data are from a survey administered to call center employees gathering health metrics for call center workers. The survey covers a time period of eight months. We will apply different methods, limiting the effect of missing data, to create statistical models to determine the effects of certain variables. We will create generalized linear models (GLMs) and a simple linear regression (SLR) model.

Introduction

In the age of computing, more people continue to find jobs that require them to sit all day. Physical activity at the workplace is not as common as it once was, which can potentially lead to unwanted health consequences. It is important to deduce which factors related to the workplace may have a role in influencing health metrics such as weight gain. Doing so may help prevent health issues among employees in the long-run and thus bolster workplace productivity.

Our goal is to detect whether factors, particularly total metabolic minutes, a measure of the intensity and duration of a person's physical activity in minutes per week, and shift time, affect weight gain for employees of a call center. In the exploratory data analysis, We will create correlation plots, density plots, and a scatterplot to visualize these relationships. We will apply our early findings to our model selection process in order to generate models that could answer our questions of focus. The models will be two GLMs: one that will include shift as a predictor and one that will not, with each having a binary Yes/No response variable of weight gain. We will also build an SLR model with a continuous numerical response variable of pounds gained. This report details our analytical decisions and insights.

The Data

The data used in this analysis were provided by the call center. The data consist of 392 observations and 83 variables, including 12 variables of interest. We remove 40 empty observations. Total metabolic minutes is a critical variable in this analysis, and some observations are missing this variable value. However, we calculate and replace many of these missing values using the following equation:

$$Total\ Metabolic\ Minutes = 8 * Vigorous\ Exercise\ Time + 4 * Moderate\ Exercise\ Time + 3.3 * Walking\ Exercise\ Time$$

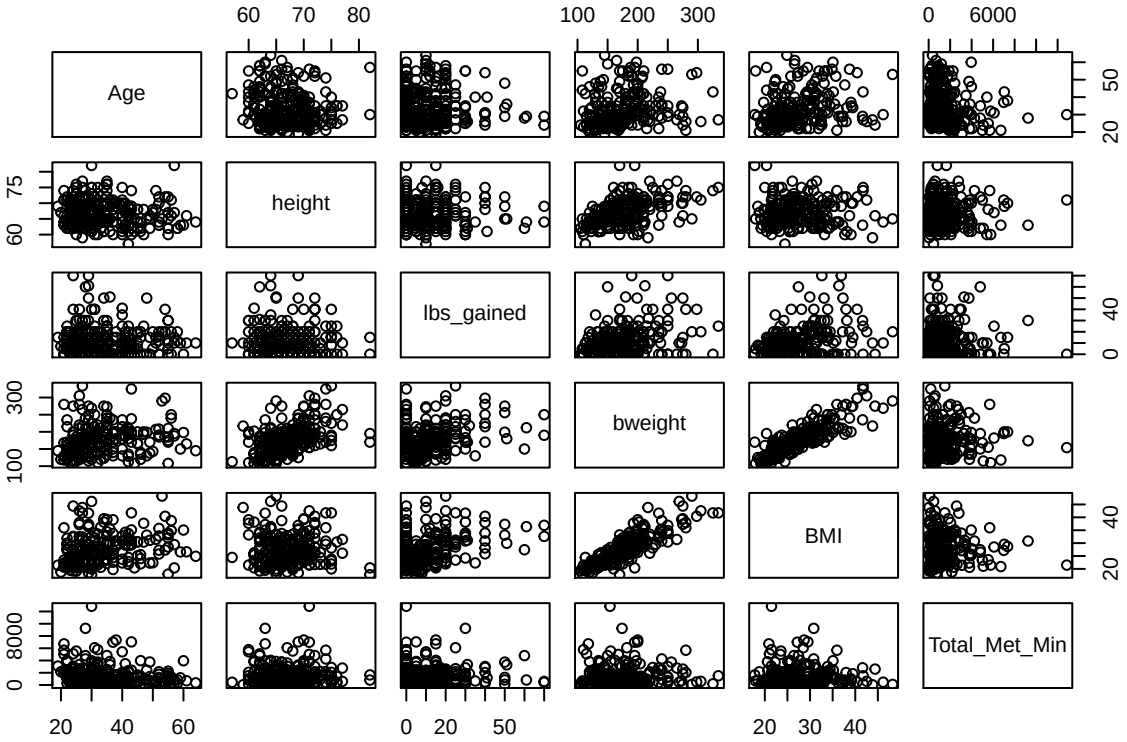
For the numerical pounds gained variable, there are many N/A values. For observations that have a binary weight gain value of "No," we change the numerical pounds gained value to 0 to indicate that this observation did not gain weight. Because the numerical pounds gained variable cannot be calculated using a transformation of other variables, we remove certain observations that have missing pounds gained and weight gain variables.

We calculate and input values for beginning weight by subracting pounds gained from body weight. We also calculate beginning BMI using a transformation of beginning weight and height:

$Beginning\,BMI\ =\ 730\ *(Beginning\,Weight\ /\ (\ Height^{\,2}))$

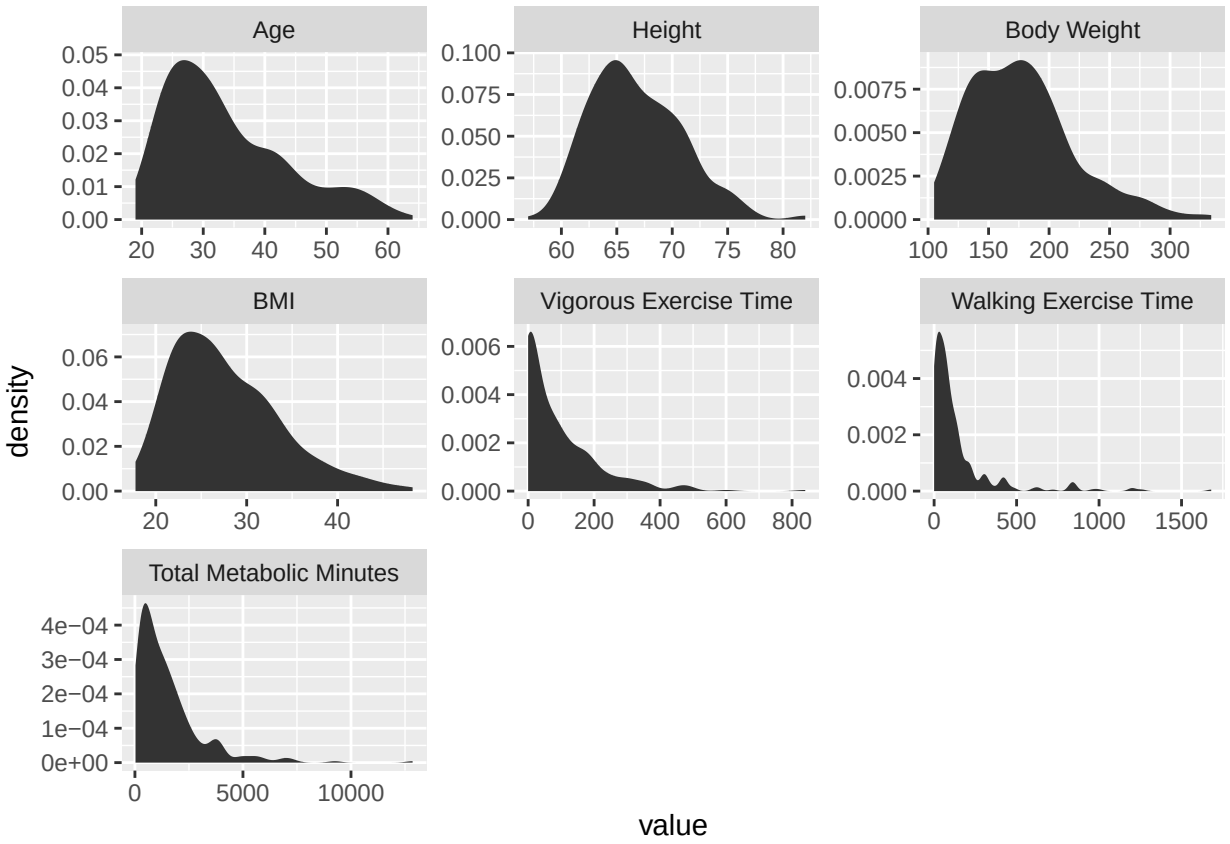
Exploratory Data Analysis

In order to see whether certain covariates are correlated, We create pairs plots to observe basic trends. We look at continuous numerical variables, because their trends are more meaningful than factor variables. We also look at the shift variable, because it is numerically ranked and is very relevant to this study's primary purpose.



As expected, body weight and BMI have a strict positive correlation, as BMI is linearly dependent upon body weight. Height also has a positive relationship with lbs_gained, as intuitively expected. Besides these and other expected relationships, there are not many easily discernible correlations. Specifically, neither shift nor total metabolic minutes has an evident relationship with pounds gained.

We now will observe density plots of certain variables to see whether any transformations will be beneficial.



Although several variables are right-skewed, We see that Age and BMI are not only right-skewed but also have no zero values, so we will do log transformations on these variables. This will aid in the creation of the linear models by making the assumptions of normality more realistic.

We will group different shift values together as follows:

Early morning: shift beginning from 7 to 9am

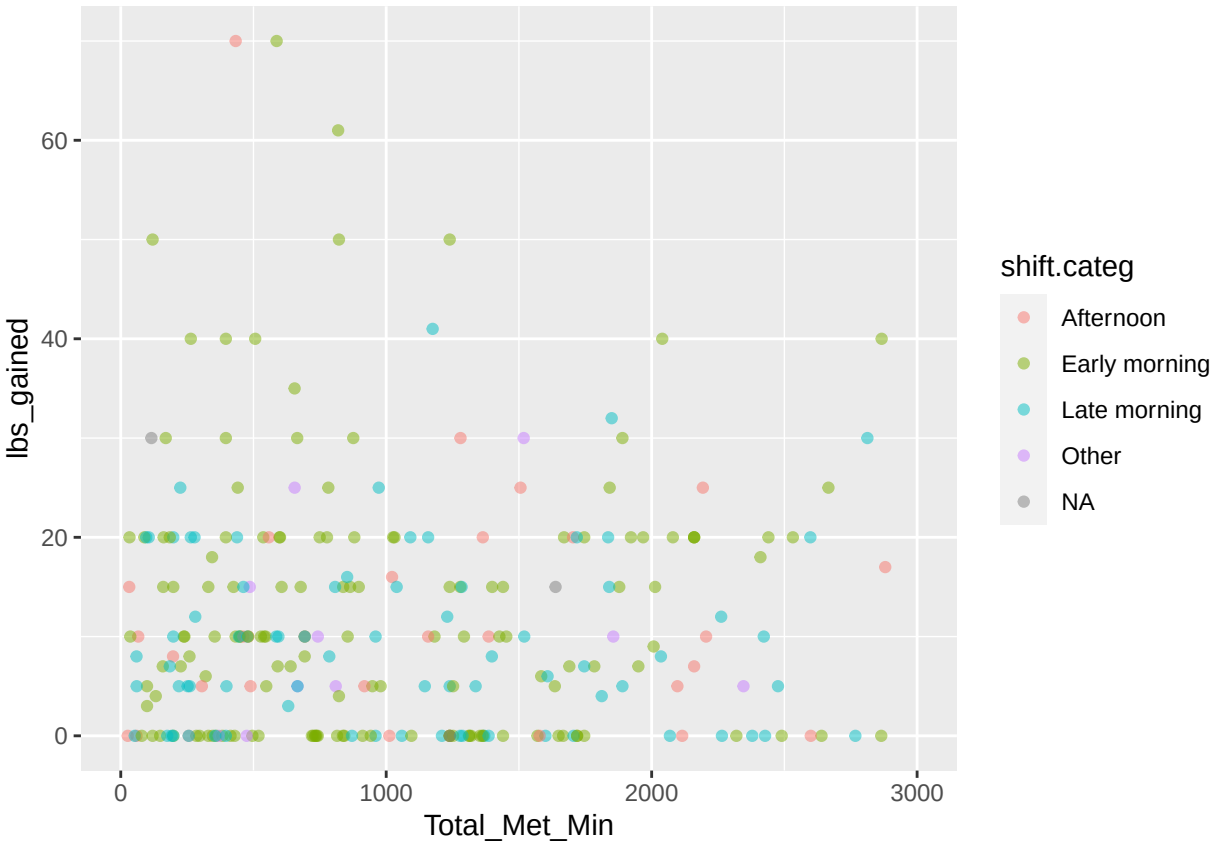
Late morning: shift beginning from 10 to 11am

Afternoon: shift beginning from 12 to 2pm

Other: other

N/A: NA

To analyze potential effects of shift and total metabolic minutes simultaneously, we look at a plot with total metabolic minutes as the independent variable and pounds gained as the response variable, with observations colored by shift category.



The response variable, pounds appears to be evenly scattered among all values of the independent variable, total metabolic minutes. This, like the earlier pairs plots, suggests that total metabolic minutes may not strongly influence pounds gained. Moreover, the different colors on the plot do not seem to follow any particular pattern or trend, suggesting, like the earlier pairs plots, that shift may not be very relevant in predicting pounds gained either.

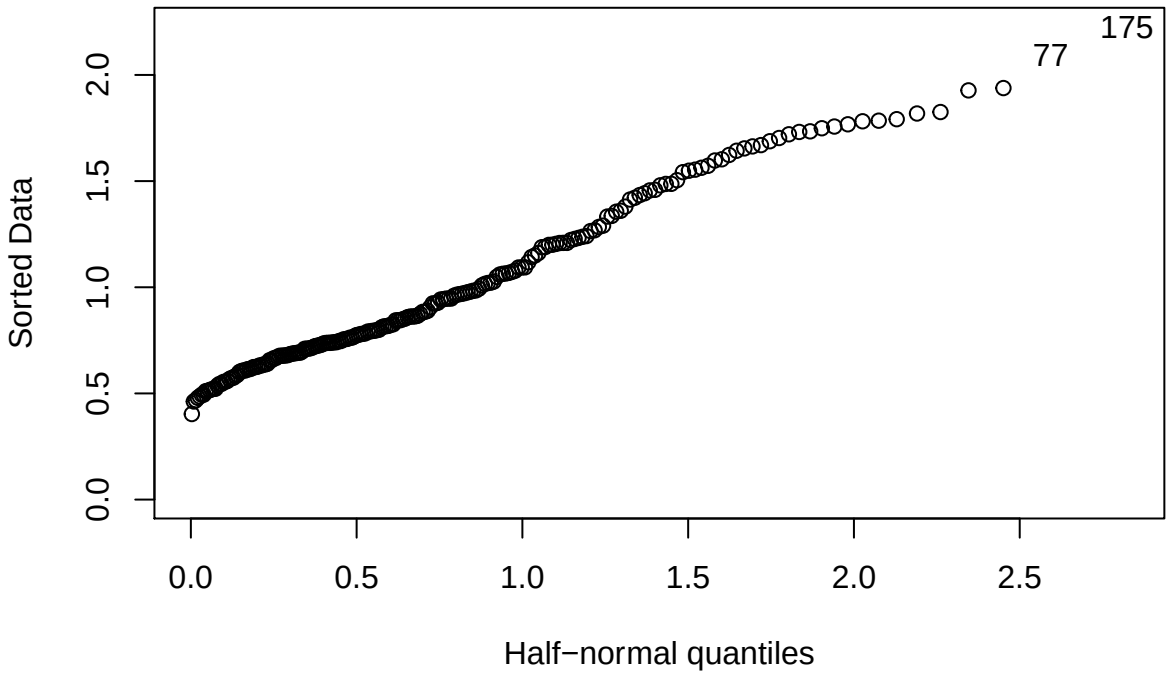
Modeling

We want to create two different models. One model will be a binary logistic regression model, which will have a Yes/No response variable of weight gain. The second model will be a simple linear regression model, with the numerical pounds gained variable as the response. There are benefits and drawbacks to each model. The generalized linear model (GLM) will have more realistic interpretations for observations that had missing numerical pounds gained values, whiole the simple linear regression model will be better at quantifying the effects of certain covariates.

Model 1: Generalized Linear Model (GLM)

By turning weight gain into a dummy variable and performing a logistic regression, we can see whether other variables have influence on weight gain. First, we factor the two categorical variables and build a model with variables including shift category, gender, age, height, and beginning BMI and total metabolic minutes. Because the change in BMI is very minimal for the vast majority of observations, we decide to use the beginning BMI as one of the regressors.

We create two GLMs, one that includes shift category as a predictor and one that omits shift category. We will then perform a Chi-squared test to see which of the two models is preferred.



The half normal plot for the GLM that includes shift category shows that there are no obvious outliers in the model.

We now look at the significance of variables in this model.

```
##                            Coefficient Estimate P-value
## Shift time: Early Morning -0.7879               0.2947
## Shift time: Late morning  -0.3766               0.6314
## Shift time: Afternoon     -1.0870               0.8240
## Gender: Female             1.6100               0.2903
## Gender: Male               0.7895               0.6010
## Beginning BMI             -0.0669               0.0205
## Total Metabolic Minutes   -0.0001               0.0890
```

In the model that includes shift category as a predictor of the binary weight gain response, only beginning BMI appears to be significantly indicative using a confidence level of 95%. None of the shift categories seem to influence weight gain.

We now look at the significance of variables in a new GLM that does *not* include shift category.

```
##                            Coefficient Estimate P-value
## Shift time: Early Morning -0.7879               0.2947
## Shift time: Late morning  -0.3766               0.6314
## Shift time: Afternoon     -1.0870               0.8240
## Gender: Female             1.6100               0.2903
## Gender: Male               0.7895               0.6010
## Beginning BMI             -0.0669               0.0205
## Total Metabolic Minutes   -0.0001               0.0890
```

With the new model not including shift category, beginning BMI again is the only significant predictor of the binary weight gain response.

We perform a Chi-sqaure test to test whether the model that contains shift category is preferred over the model without shift category. The null hypothesis is that the reduced model, i.e.: the model without shift category, is a superior model, and the alternative hypothesis is that the model without shift category is not

a weaker 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.

```
##          P-Value
## Chi-test 0.635
```

We get a P-value of 0.635, which is very large given a confidence level of 95%. Therefore, we fail to reject the null hypothesis that the reduced model is superior. In other words, there is not enough evidence to favor the model that includes shift category. It is evident that shift is not significantly indicative of weight gain.

Model 2: Simple Linear Regression (SLR)

We have seen that beginning BMI is a significant indicator of binary weight gain, and that shift category is not. We now want to perform similar analysis, although with a continuous numerical variable, pounds gained, as the response variable. To do this, we will use a simple linear regression (SLR) model. We will not include any of the shift categories in this model, as the previous model indicated that no shift group is a signficant predictor.

```
##                          Coefficient Estimate P-value
## Gender: Female           8.8372               0.4970
## Gender: Male             3.3178               0.8010
## Age                      2.9601               0.4290
## Height                   0.5737               0.0880
## Beginning BMI            0.0233               0.9050
## Total Metabolic Minutes 0.0294                0.1560
```

Using a confidence level of 95%, no predictor in this model is significantly indicative of a certain amount of pounds gained. For both this SLR model and our earlier preferred GLM, total metabolic minutes was not a significant predictor.

Discussion and Conclusion

The Relationship Between Total Metabolic Minutes and Weight Gain

The purpose of this analysis was to draw conclusions from call center employee health data, specifically regarding the effect of total metabolic minutes and shift time on weight gain. From our exploratory data analysis, it appeared from the beginning that shift time might not be a particularly useful variable in predicting numerical pounds gained. Our GLMs confirmed this, with one GLM indicating that no specific shift group was statistically significant, and the other reduced GLM without shift performing better than the former GLM in a Chi-squared test.

Becuase total metabolic minutes is a continuous numerical variable, we decided to use a model building process that was not as focused on answering a binary Yes/No question regarding weight gain. We instead created a simple linear regression model to test the significance of numerical variables and gender. This model indicated that total metabolic minutes also is not a statistically significant predictor for the numerical pounds gained variable.

Although it is unfortunate that we could not reach many conclusive results on the effectiveness of total metabolic minutes and shift time as potential indicators of binary and numerical weight gain, we conclude this analysis knowing that there is a path to discover more. We could ascertain from certain areas of our model summaries that gender and beginning BMI might be useful indicators of weight gain. It would be worth conducting future analyses on these specific variables to see whether any conclusions could be made on these variables' effects.