The Relationship Between Exercise Time and Weight Loss

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 from a start time for call center workers. The survey covers a time period of eight months. We will apply different methods to limit the effect of missing data and to create statistical models to determine the effects of certain variables. We will create *insert models here, describing response variable for each*

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.

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. We will create *descriptive statistics and plots* to visualize these relationships in order to generate models that could answer these questions. The models will be *insert models and response variables*. 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:

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

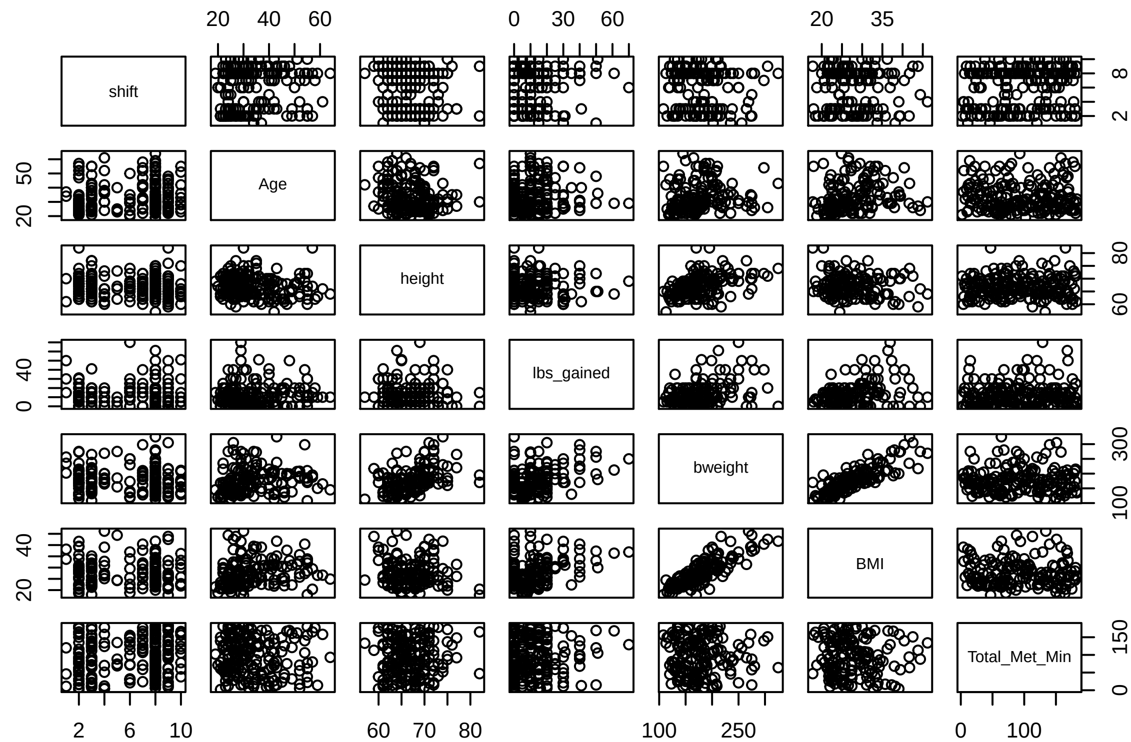
For the numerical pounds gained variable, there are many N/A values. For observations that have a binary weight gain value of "No," we change the numerical pounds gained value to 0 to indicate that this observation did not gain weight. Because the numerical pounds gained variable cannot be calculated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

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

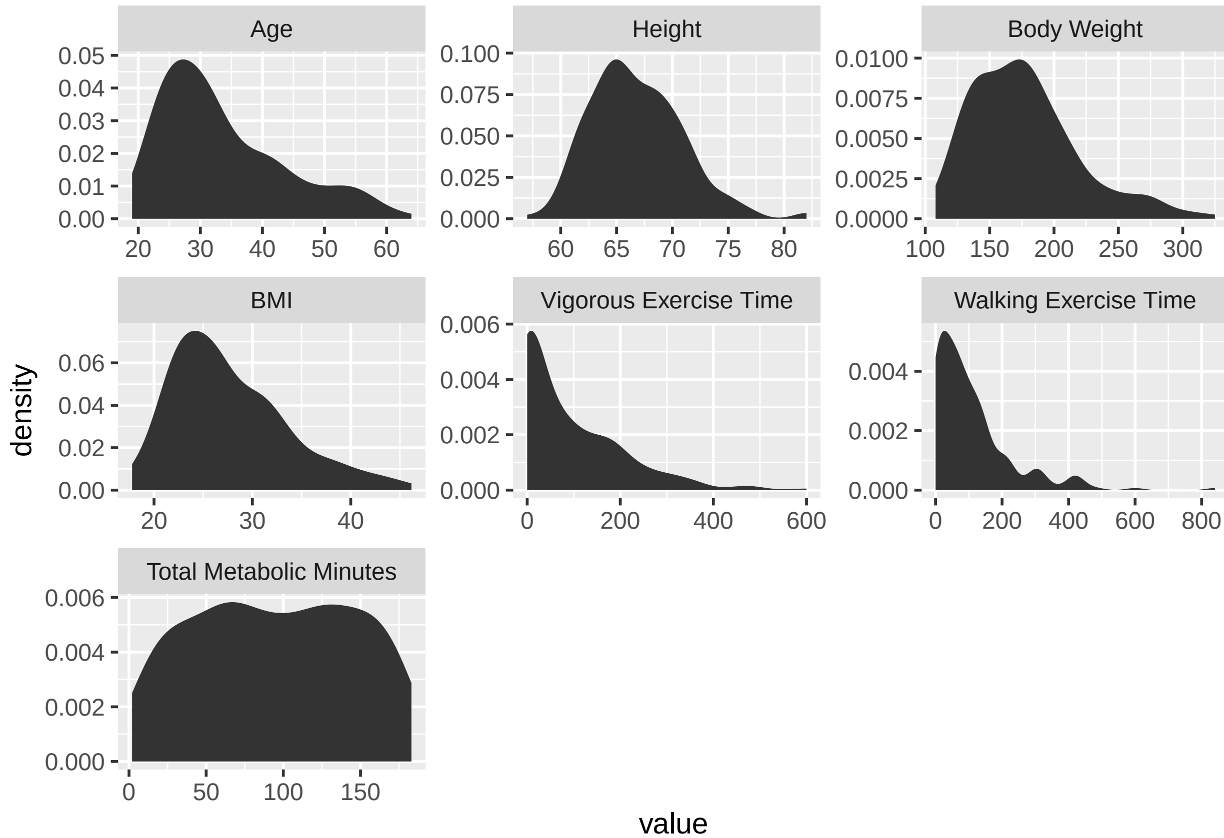
Exploratory Data Analysis

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

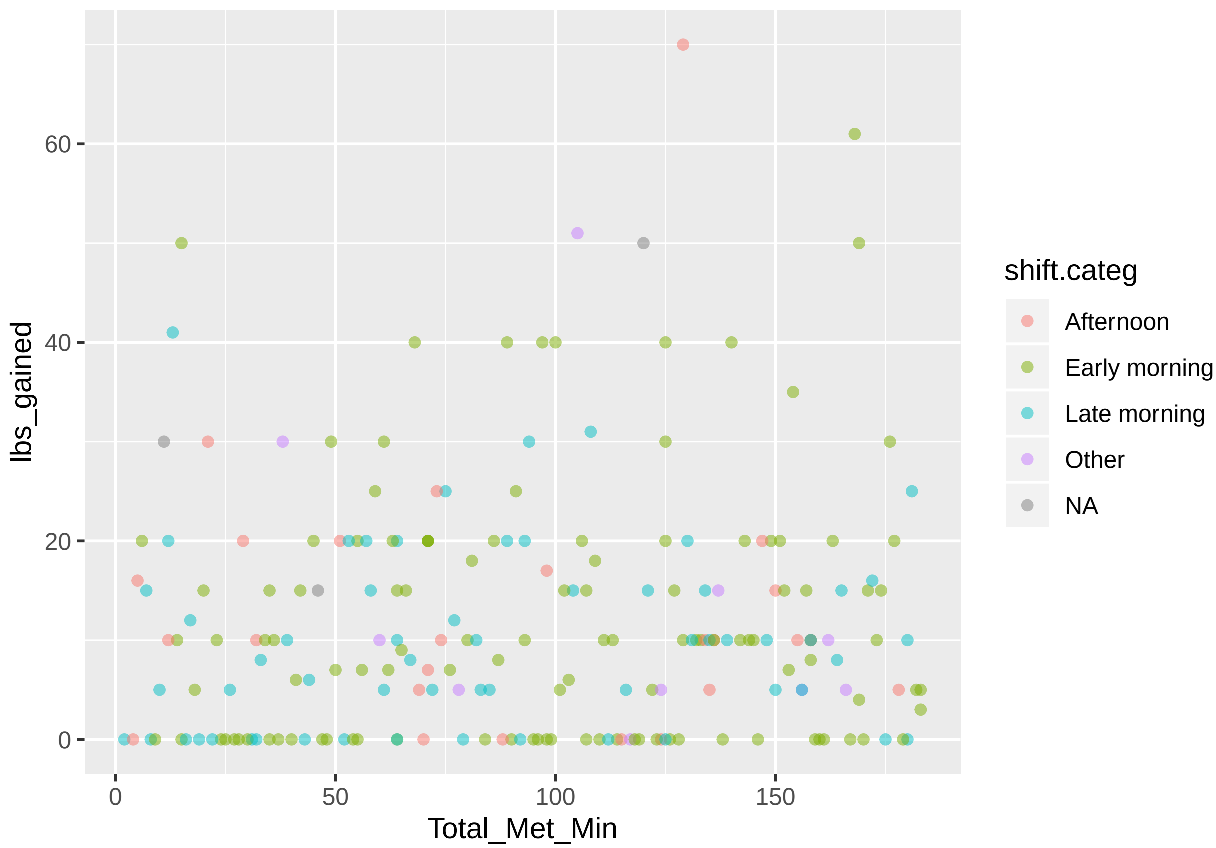
Late morning: shift beginning from 10 to 11am

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 scatterplot with total metabolic minutes as the independent variable and pounds gained as the response variable, with observations colored by shift category.



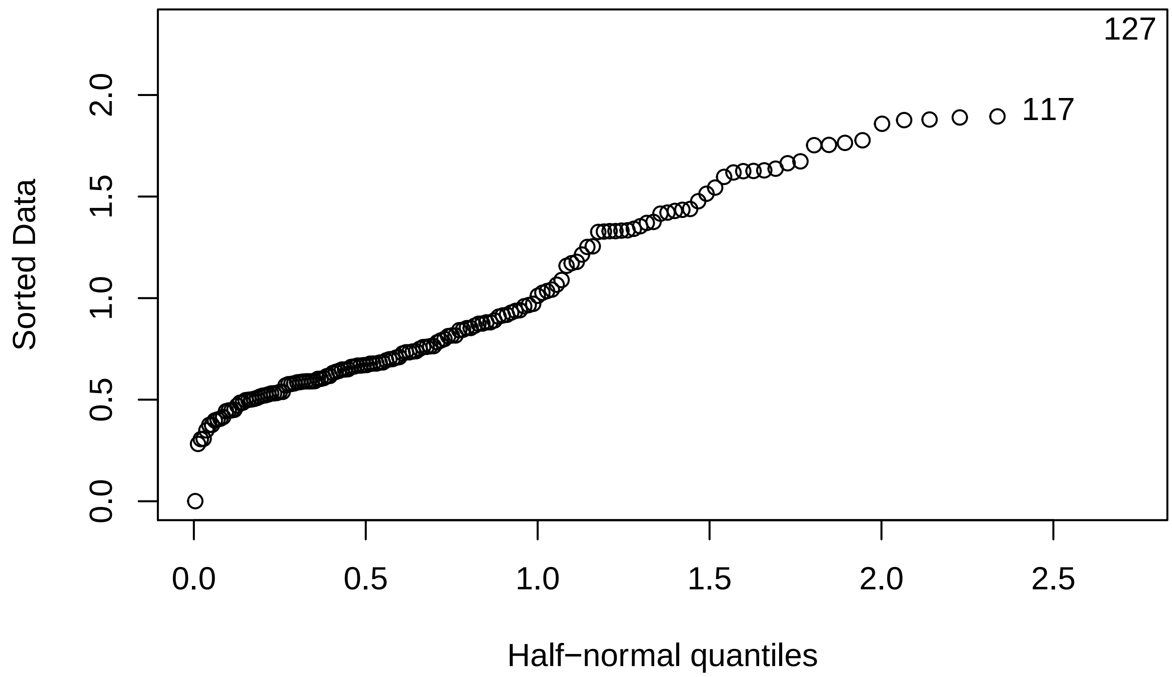
The response variable, pounds appears to be evenly scattered among all values of the independent variable, total metabolic minutes. This, like the earlier pairs plots, suggests that total metabolic minutes may not strongly influence pounds gained. Moreover, the different colors on the plot do not seem 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.



The half normal plot shows that there are no obvious outliers in the model.

When weight gain is used as a binary variable, taking on the values yes or no, we can use logistic regression to measure the effect of our explanatory variables on whether an observation did or did not gain weight. For this logistic regression model, shift time, gender, age, height, beginning BMI, and total metabolic minutes are the explanatory variables.

```
##                            Coefficient Estimate P-value
## Shift time: Early Morning             -0.6210  0.2796
## Shift time: Late morning              -0.3274  0.5895
## Shift time: Afternoon                 -1.0870  0.8240
## Gender: Female                         1.6100  0.2903
## Gender: Male                           0.7895  0.6010
## Beginning BMI                         -0.0670  0.0205
## Total Metabolic Minutes               -0.0001  0.0890
```

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.

```
## Analysis of Deviance Table
```

```
##
## Model 1: weightgain.b ~ gender + Age + height + Beg_BMI + Total_Met_Min
## Model 2: weightgain.b ~ shift.categ + gender + Age + height + Beg_BMI +
##     Total_Met_Min
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       147     159.65
## 2       144     157.94  3   1.7088    0.635
```

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.

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.

Discussion and Conclusion

ascertain

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.