SAP

Group 1: Brendan Winters, Brandon Kill, Paul Hunt

9/14/2021

Introduction

One may posit that the rapid development of, and increase in, computer-related jobs has led to a decline in individual activity over the past century. Some companies are even incurring additional healthcare costs, as employees are more sedative at work than ever before in recorded history. As a response, many large employers have installed fitness centers on their campus, and one has enlisted our help analyzing some fitness data they gathered from their employees (submitted voluntarily). Our task is to determine if total metabolic minutes or an employee's shift start time affect whether and how much weight they gain in an eight-month period.

Analysis Population

This study will draw upon a data set consisting of several health metrics self-reported within an eight month period by a call center's employees. The data set contains metrics for 392 employees, but only 342 observations include the response pounds gained. Table 1 shows summaries of the responses to various questions among those who gave the number of pounds they gained or indicated that they did not gain weight (recorded in the data as 0 lbs. gained). Some potential predictors included in the study are gender, age, height, weight, department, job, and shift time, which is one of the two main predictors that the employer is most interested in.

Another predictor variable given is exercise time, which is broken down to vigorous exercise time, moderate exercise time, and walk exercise time, all measured in minutes. These three exercise times form the composite metric "Total Metabolic Minutes," which assigns weights of 8, 4, and 3.3 to the three categories of exercise times, respectively. Total Metabolic Minutes is the second main predictor of weight gain that the employer is most interest in.

Missing observations in the predictors height, age, and job may prove to be a problem. These may need to be imputed before proceeding with analysis. Most of the missing observations for total metabolic minutes could by calculated with the three exercise times given, and these are the values shown in Table 1.

For the response, we have an indication as to whether or not employees gained weight as well as the number of pounds they gained. Finally, we have BMI and base weight at our disposal, which may be of use to us when detecting potential outliers.

Table One: Variables of Interest and Summary Statistics

Variable	**Count (%)**	**n Missing**	**Mean** (S.D.)
lbs. Gained	342	0	11.37 (12.99)
Gender	337	5	
male	98 (29.08)		
female	239 (70.92)		
Shift	338	4	
7am	28 (8.28)		
8am	112 (33.14)		
9am	55 (16.27)		
10am	49 (14.5)		
11am	42 (12.43)		
12pm	14 (4.14)		
1pm	8 (2.37)		
2pm	15 (4.44)		
other	15 (4.44)		
Job	320	22	
611	34 (10.62)		
AOC	26 (8.13)		
Collections	68 (21.25)		
COOS	67 (20.94)		
Employee Accounts	4 (1.25)		
Executive Relations	3 (0.94)		
FRC	14 (4.37)		
Internal	1 (0.31)		
OCA	10 (3.12)		
Operations	6 (1.87)		
Refunds	3 (0.94)		
Resource Mgt.	9 (2.81)		
Response	2 (0.63)		
Tech Support	12 (3.75)		
other	61 (19.06)		
Department	335	7	
CFS	171 (53.44)		
CS	108 (33.75)		
Facilities	3 (0.94)		
HR	5 (1.56)		
IT	6 (1.87)		
Training	10 (3.12)		
Other	32 (10)		
Age	314	28	33.75 (9.95)
Height (in.)	322	20	66.67 (4.07)
Base Weight	264	78	177.75 (43.94)
BMI	249	93	27.87 (6.1)
Vigorous Excercise Time	342	0	77.76 (115.11)
Moderate Excercise Time	341	1	74.99 (142)
Walk Excercise Time	342	0	124.83 (213.37)
Total Metabolic Minutes	341	1	1329.37 (1567.88)

Specific Aims:

1) Determine if total metabolic minutes have an effect on weight gain.

1.1 - The histogram of lbs_gained (Figure 2) indicates that we are dealing with a truncated model. The questionnaire asked if the subject had gained weight, and if so, how many pounds they gained. The questionnaire did NOT give an opportunity for those who lost weight to report their negative weight change. So, our first step (after removing the subjects who failed to fully complete the survey) will be to separate those who gained weight from those who did not.

1.2 - Further exploratory data analysis will reveal whether any variable mutations (percentages, differences, etc.) or transformations (log, polynomial, etc.) will be necessary.

1.3 - We will then consider fitting a binomial logit model for the binary variable weightgain ("Yes" or "No"), and we will use metabolic minutes (as well as the rest of the relevant available predictors) to see if we can reliably model the likelihood of whether or not a subject gains weight.

1.4 - Afterwards, we will attempt to use zero-inflated poisson regression to model the number of pounds gained, again using total metabolic minutes and the rest of the relevant predictors. This is a piecewise regression process that includes first, the binomial model for gaining weight or not, and then a Poisson model for the number of pounds gained.

1.5 - Given the shape of the histogram of weight gain, and especially the wording and design of the survey, other viable tools such as truncated or censored regression analysis could also be considered when determining if total metabolic minutes affects weight gain.

2) Does shift have an effect on weight gain?

2.1 - To analyze the impact an employee's shift has on weight gain, we will initially use nested mixed effects to account for the hierarchical nature of the categorical data in our weight gain model. One's shift is inherently nested inside one's job which is inherently nested inside one's department, encouraging us to use multilevel regression analysis in this case.

2.2 - Because our response is not Gaussian (Figure 3) in nature, we will need to use generalized mixed effects regression so that we can utilize the different link functions to the binomial (For whether or not weight was gained) and poisson (for modeling the number of pounds gained) distributions.

2.3 - The next step is to use these results to see if we need to make any additional, yet similar models, with different combinations of the nested variable hierarchy in (2.1) as random effects and fixed effects.

2.4 - We will then use the Restricted Maximum Likelihood approach for determining goodness-of-fit for the mixed effect models in order for us to best determine if shift affects weight gain.

2.5 - If we run into issues with fit, Canonical Correlation Analysis (CCA) could be used to identify multicolinearity between an employee's shift, job, and department. We would not be surprised if having the same job correlated to being in the same department and/or having the same shift. (See Figure 1 for scatterplot matrix)

2.6 - If truncated or censored regression winds up being the superior model (1.5), we will still attempt to first utilize nested mixed effects to deal with shift, job, and department, and then we will adjust accordingly if need be in similar ways as those stated above.

Conclusion

When it comes to analyzing the data provided, the lack of opportunity for subjects to indicate the existence and quantity of weight loss begets a very specific challenge, as an employee's change in weight is now left-truncated at zero pounds. We have many tools at our disposal to handle this phenomenon, and we will utilize the statistical methods outlined above to provide you with meaningful answers as to whether and to what extent metabolic minutes and an employee's shift impact weight gain.

Supplamentary Figures

Fig. 1: Scatterplot Matrix of Variables of Interest

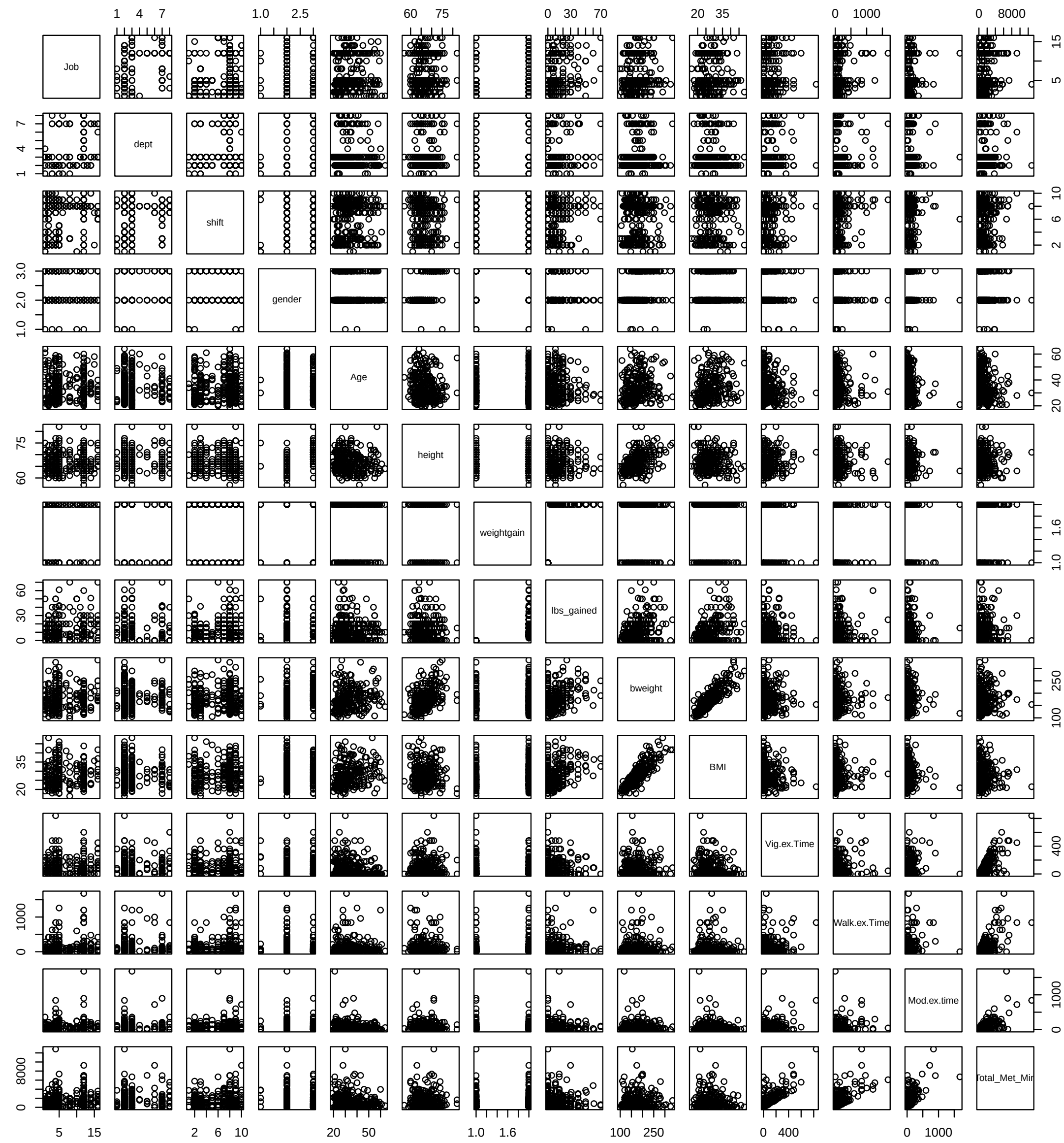

Figure 1: The above scotterplot matrix shows the scatterplots for relationships between each of the variables considered in this analysis. It does not seem to indicate that covariance between the predictors will be an issue. However, BMI is calculated as a function of height and weight, and total metabolic minutes is calculated as a function of the three exercise time variables, so we must take that into consideration when building our model. Depending upon which variables we end up using, BMI and exercise time may still wind up being used in the model.

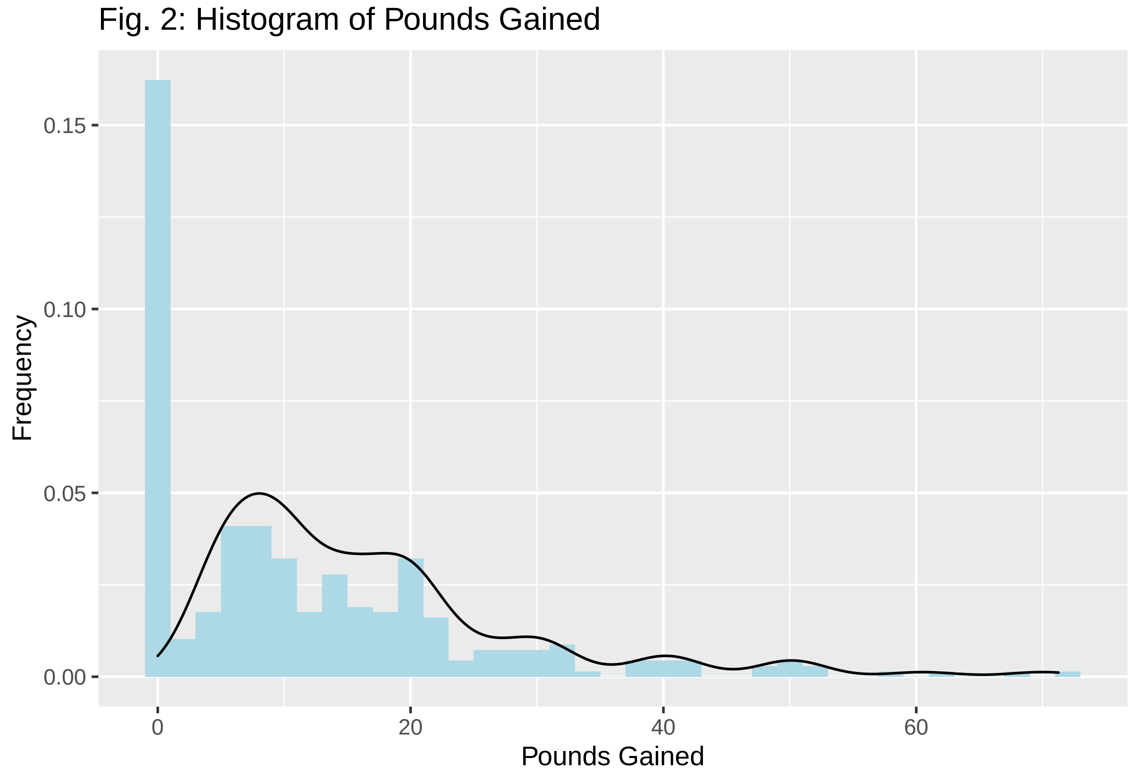

Figure 2: The above histogram shows the relative frequencies of pounds gained for employees at the call center, while the overlayed density curve shows the emperical density curve of pounds gained amongst emplyees who reported gaining weight. The pounds gained have been uniformly jittered for the non-zero responses to smooth some peaks at multiples of 5, likely due to respondants rounding their weight gain. The high proportion of respondents who did not report gaining weight indicates that a higherarchical model differentiating between those who gained weight and those who did not may be appropriate. The long right tail of the density curve indicates that a Poisson GLM may be appropriate to model the pounds gained amongst those who reported weight gain.

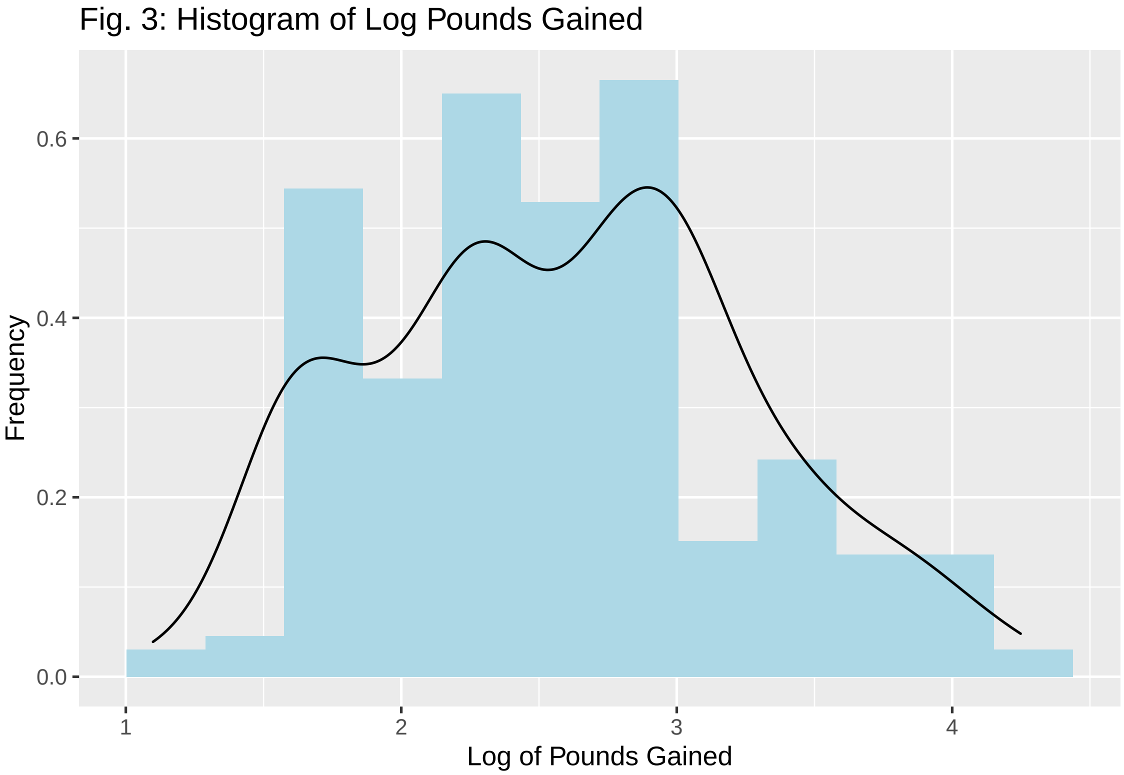

Figure 3: The above histogram shows the natural log of pounds gained amongst those who reported weight gain with the continuous density estimate overlayed. The relatively fat left tail and skinny right tail suggest that a log transformation may not be sufficient to create normally distributed residuals in a normal LM for the weight gained level of a hierarchical model.