SAP

Group 1: Brendan Winters, Brandon Kill, Paul Hunt

9/10/2021

Introduction

One may posit that the rapid development of, and increase in, computer-related jobs has led to a substantial decline in individual activity over the past century. The ever-deepening penetration of technology into our lives certainly seems likely to be a major contributor. As result, it is becoming more and more common for workers to become overweight or to incur an obesity-related disease which may be of considerable cost to a company. This analysis intends to determine which factors may affect weight gain so we might develop a better understanding of how to reduce such costs surrounding obesity in the workplace.

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. We will attempt to determine which of these metrics affects weight change over that period, and to what degree. The data set contains metrics for 392 employees, but only 342 observations include the response pounds gained. Table one shows summaries of the responses to various questions amongst 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).

The potential predictors included in the study are gender, age, height, weight, department, job, and shift time. As well as exercise time broken up by vigorous exercise, moderate exercise, and walk exercise all measured in minutes. These exercise also form a composite metric "Total Metabolic Minutes," which weights more vigorous exercise more highly.

For the response, we have an indication of whether or not employees gained weight, and how many pounds they gained. In addition, we have BMI and base weight at our disposal, which may be of use to us – especially in the analysis of 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)
b 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	247	95	1124.33 (1100.02)

Specific Aims:

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

- 1.1 The histogram of pounds gained 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.
- 1.5 Given the shape of the histogram of weight gain, other viable models could also be considered when determining if, how, and to what extent 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 use a linear mixed effects model to account for the hierarchical nature of the categorical data. One's shift is inherently nested inside one's job which is inherently nested inside one's department, which encourages us to use multilevel regression analysis in this case.
- 2.2 The next step is to make multiple models, choosing different combinations of the variable hierarchy in (2.1) to be fixed effects and nested random effects.
- 2.3 We will then use the Restricted Maximum Likelihood approach for determining goodness-of-fit for the mixed models in order for us to best determine if, how, and to what extend shift affects weight gain.

Conclusion

When it comes to the data provided, the lack of opportunity for subjects to indicate the existence and quantity of weight loss exhibits a troubling lack of foresight and planning in the data collection process, and the sheer amount of partially-completed surveys displays a stunning lack of supervision for which we are not responsible. However, we are confident that we can clean up the data provided in such a way that it can still be useful in our analysis. We will utilize the statistical methods outlined above to provide you with meaningful answers as to whether metabolic minutes or an employee's shift have an impact on weight gain, and if so, in what way, and by how much.

Fig. 1: Scatterplot Matrix of Variables of Interest

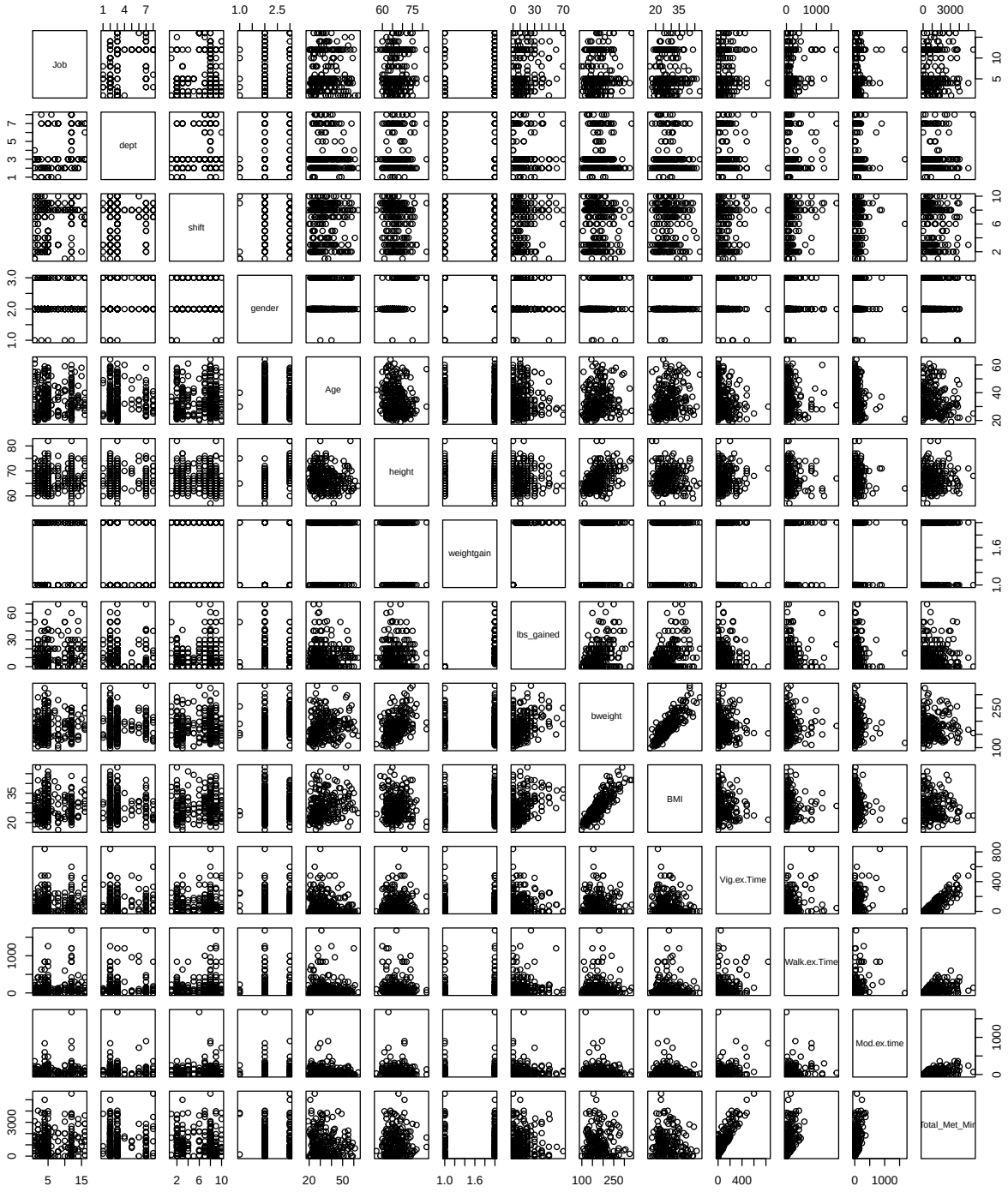

Figure 1: The above scotterplot matrix shows the scatterplots for relationships between each of the variable considered in this analysis. It does not seem to indicate that covariance between the predictors will be an issue, except for between BMI, and height and weight, and between Total metabolic minutes and the three variables for excercise time. Thus BMI and the excercise times may need to be excluded.

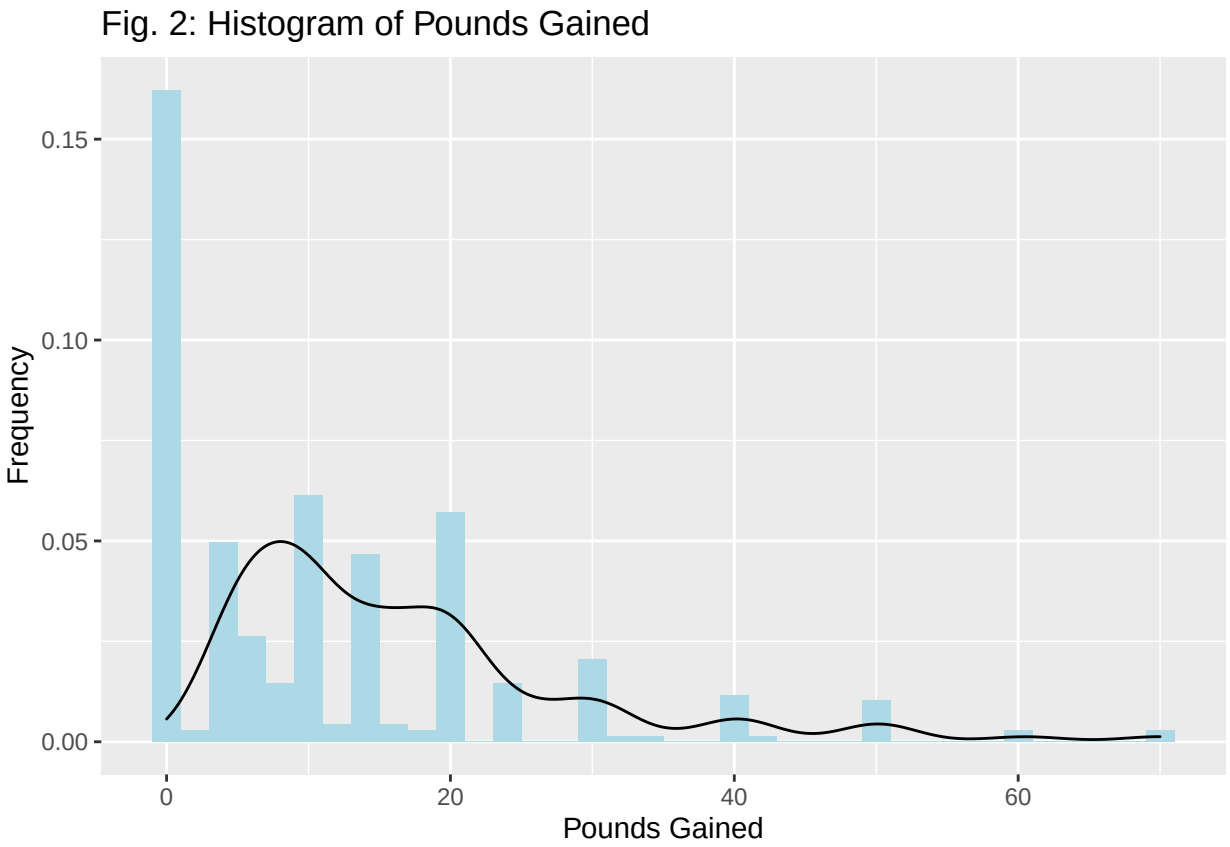

Figure 2: The above histogram shows the relative frequencies of pounds gained for employees at the call center, while the overlayed density curve shows the emperical density curve of pounds gained amongst emplyees who reported gaining weight. The high proportion of respondents who did not report gaining weight indicates that a higherarchical model diffenetiating 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.

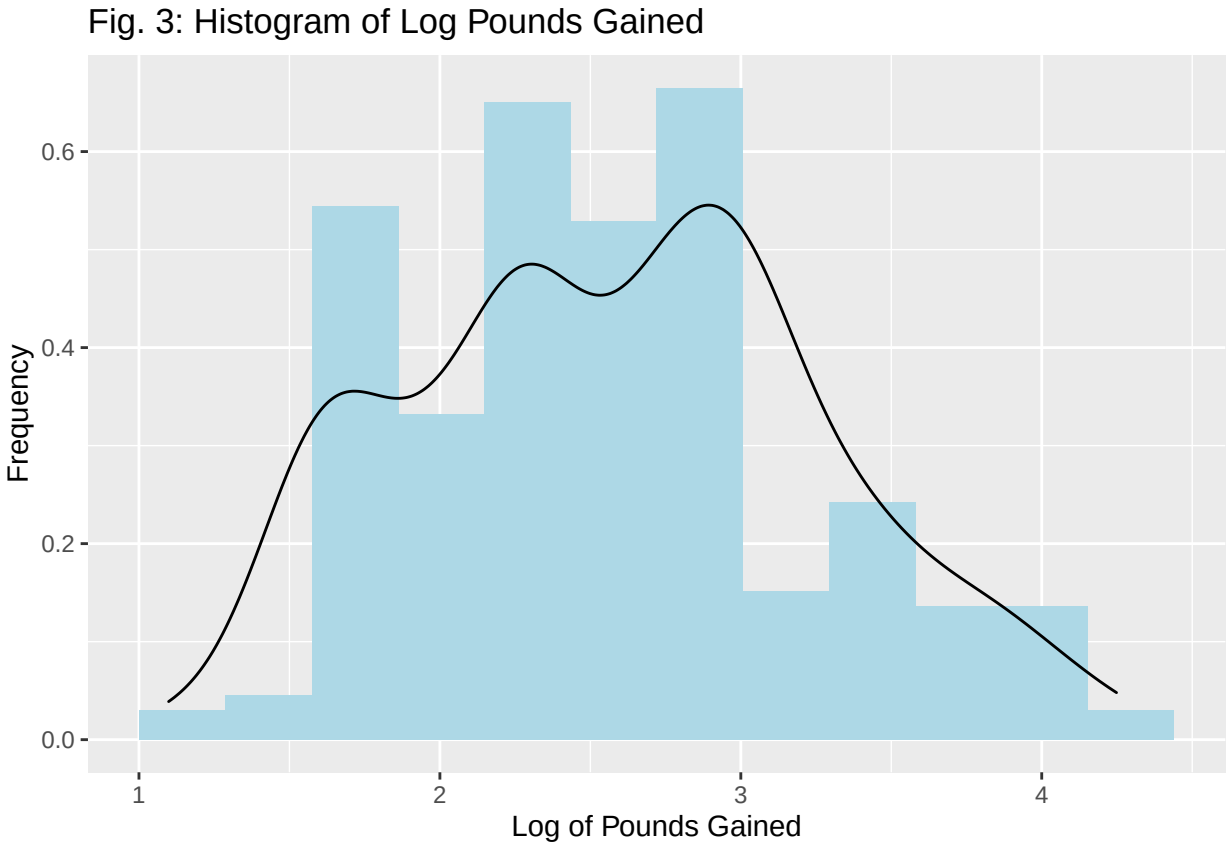

Figure 3: The above histogram shows the natural log of pounds gained amongst those who reported weight gain with the continuous density estimate overlayed. The relatively fat left tail and skinny right tail suggest that a log transformation may not be sufficient to create normally distributed residuals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